

**Análisis espaciotemporal de factores de riesgo de deforestación en los parques nacionales
Chibiriquete, Tinigua y sierra de la Macarena utilizando datos de puntos de calor,
cicatrices de quema y explotación ganadera**

Diego Alejandro Estrada Vélez

Asesor

Felipe Alexander Pipicano Guzmán

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI
Especialización en Ciencia de Datos y Analítica

2024

Dedicatoria

A mi familia por su comprensión y por su apoyo.

Agradecimiento

A la vida por la gran oportunidad de realizar y completar este nuevo proyecto de formación y crecimiento.

A la Universidad Nacional Abierta y a Distancia (UNAD) por brindar este programa que permite que muchas personas de diferentes condiciones y geografías tengan acceso a este nivel de formación.

Al profesor Felipe Alexander Pipicano Guzmán por su orientación durante el desarrollo del presente trabajo.

Resumen

Este trabajo de grado se enfoca en el análisis espaciotemporal de los factores de riesgo de deforestación en los Parques Nacionales Chibiriquete, Tinigua y Sierra de la Macarena, ubicados en la Amazonía colombiana. Mediante la aplicación de modelos estadísticos y de machine learning, tales como la regresión lineal múltiple, el modelo aditivo generalizado (GAM) y la regularización Ridge, se desarrolla un Indicador de Riesgo de Deforestación (IRD) que facilita la identificación y priorización de áreas con alto riesgo de deforestación. Este indicador se fundamenta en variables como puntos de calor, cicatrices de quema y explotación ganadera, seleccionadas por su relevancia en la dinámica de deforestación en la región. Se verifica que los datos disponibles son suficientes, pertinentes y adecuados para la construcción de modelos efectivos. Además, se emplean herramientas como Google Earth Engine y Dynamic World para superar las limitaciones asociadas a la falta de actualización de algunas variables de riesgo, permitiendo desarrollar modelos de estudio pertinentes y actuales. Para garantizar la robustez y capacidad de generalización de los modelos, se llevan a cabo validaciones cruzadas y ajustes de hiperparámetros, evaluando diversas configuraciones para optimizar el rendimiento y mitigar el riesgo de sobreajuste. Los resultados reflejan un buen desempeño general de los modelos, aunque se observan ligeras diferencias entre los errores en la validación cruzada y los obtenidos en el conjunto de prueba. A pesar de estas discrepancias, ambos modelos demuestran una capacidad adecuada para predecir la degradación del bosque, lo cual justifica su uso en la generación del IRD. Las recomendaciones derivadas de este estudio incluyen la incorporación de variables adicionales como densidad poblacional, cultivos lícitos e ilícitos, y la infraestructura de vías terciarias, así como el uso de tecnologías de monitoreo remoto, como Google Earth y Dynamic World, para mejorar la precisión y actualización de los datos. Asimismo, se resalta la

importancia de validar el IRD mediante comparaciones con otras fuentes reconocidas, como Global Forest Watch, y de incorporar retroalimentación de expertos y actores locales para asegurar la utilidad y relevancia del indicador en la toma de decisiones relacionadas con la conservación.

Palabras claves: Deforestación, amazonía, factores de riesgo, indicadores, regresión múltiple.

Abstract

This work focuses on the spatiotemporal analysis of deforestation risk factors in the Chibiriquete, Tinigua, and Sierra de la Macarena National Parks, located in the Colombian Amazon. By applying statistical and machine learning models, such as the generalized additive model (GAM) and Ridge Regularization, a deforestation risk indicator (DRI) is developed to facilitate the identification and prioritization of areas at high risk of deforestation. This indicator is based on variables such as fire hotspots, burn scars, and livestock activities, selected for their relevance in the deforestation dynamics of the region. It has been verified that the existing data are sufficient, relevant, and valid for constructing effective models. Additionally, tools such as Google Earth Engine and Dynamic World were employed to overcome the limitations associated with the lack of updated risk variables, allowing for the development of pertinent and up-to-date models. To ensure the robustness and generalizability of the models, cross-validations and hyperparameter tuning are performed, evaluating various configurations to optimize performance and mitigate the risk of overfitting. The results show good overall performance of the models, although slight differences were observed between the errors in cross-validation and those obtained in the test set. Despite these discrepancies, both models demonstrate adequate predictive capacity for forest degradation, justifying their use in generating the DRI. Recommendations from this study include incorporating additional variables such as population density, legal and illegal crops, and tertiary road infrastructure, as well as the use of remote monitoring technologies, such as Google Earth and Dynamic World, to improve data accuracy and timeliness. Furthermore, the importance of validating the DRI through comparisons with other recognized sources, such as Global Forest Watch, and incorporating feedback from experts and local stakeholders is

emphasized to ensure the indicator's utility and relevance in conservation-related decision-making.

Keywords: Deforestation, amazon, risk factors, risk indicators, multiple regression.

Tabla de Contenido

Introducción	15
Planteamiento del Problema	16
Justificación	18
Objetivos	21
Objetivo General	21
Objetivos Específicos.....	21
Marco Conceptual.....	22
Causas de la Deforestación	22
Consecuencia de la Deforestación	22
Métodos y Variables para Medir la Deforestación en Colombia	23
Marco Teórico.....	29
Ciencia de Datos en el Contexto de la Deforestación.....	29
Estadística Descriptiva	29
Análisis Temporal	30
Modelos de Relación entre Variables - Regresión Lineal Múltiple	30
Modelos Generalizados Aditivos (GAM)	30
Análisis de Residuos y Validación de Modelos	31
Pertinencia Actual de los Datos.....	31
Metodología	33
Pasos Metodológicos.....	33

Definición de las Áreas de Análisis	33
Procesamiento, Exploración y Análisis de Datos.....	37
Formulación de Indicadores de Riesgo	39
Descripción Espacial y Temporal de los Factores de Riesgo.....	40
Puntos de Calor 2017-2024.....	40
Cicatrices de Quema 2017-2024	43
Tendencias de Puntos de Calor dentro de PNN	46
Tendencias de Cicatrice de Quema dentro de PNN	46
Tendencia Cabezas de Ganado.....	47
Extracción, Transformación y Carga de Datos	49
Puntos de Calor	50
Cicatrices de Quema	51
Explotación Ganadera	51
Praderización en PNN.....	52
Dataset Consolidado.....	53
Análisis Exploratorio de Datos	55
Distribución de la Degradación del Bosque PNN Chibiriquete.....	55
Distribución de los Puntos de Calor PNN Chibiriquete.....	55
Distribución de las Cicatrices de Quema PNN Chibiriquete	57
Distribución de las Cabezas de Ganado PNN Chibiriquete	57

	10
Análisis de Boxplots PNN Chibiriquete	57
Distribución de la Degradación del Bosque PNN Tinigua-Macarena	60
Distribución de las Cicatrices de Quema PNN Tinigua-Macarena.....	61
Distribución de las Cabezas de Ganado PNN Tinigua-Macarena.....	61
Análisis de Boxplots PNN Tinigua-Macarena	61
Relaciones entre Variables para el PNN Chibiriquete	64
Transformación Logarítmica y Escalado	64
Modelo de Regresión Lineal Múltiple	66
Coeficientes del Modelo	66
Análisis de Correlación	69
Modelo Aditivo Generalizado	70
Coeficientes Suavizados.....	70
Evaluación de los Resultados (Validación Cruzada).....	72
Relaciones entre Variables para el PNN Tinigua-Macarena	77
Modelo de Regresión Lineal Múltiple	77
Coeficientes del Modelo	78
Análisis de Correlación	80
Regularización Ridge	82
Evaluación de los Resultados (Validación Cruzada).....	84
Definición de un Indicador de Riesgo de Deforestación para el PNN Chibiriquete	87

Propuesta de Ponderación de Pesos Usando F-value y p-value.....	88
Definición de un Indicador de Riesgo de Deforestación para el PNN Tinigua-Macarena	93
Adaptación del Método de Pesos Propuesto	93
Conclusiones	96
Recomendaciones	99
Referencias Bibliográficas	102

Lista de Tablas

Tabla 1 <i>Comparación Tecnologías de Monitoreo del Terreno</i>	27
Tabla 2 <i>Ejemplo Cálculo IRD parque Chibiriquete</i>	91
Tabla 3 <i>Ejemplo Cálculo IRD parque Tinigua-Macarena</i>	95

Lista de Figuras

Figura 1 <i>Delimitación de la Amazonía Colombiana</i>	34
Figura 2 <i>Parques Nacionales Naturales</i>	35
Figura 3 <i>Municipios en PNN Chibiriquete, Tinigua y la Macarena</i>	36
Figura 4 <i>Comportamiento Temporal Puntos de Calor en PNN</i>	41
Figura 5 <i>Boxplots Distancia Puntos de Calor a PNN</i>	42
Figura 6 <i>Comportamiento Temporal Cicatrices de Quema en PNN</i>	44
Figura 7 <i>Boxplots Distancias de Cicatrices a PNN</i>	45
Figura 8 <i>Tendencia Puntos de Calor 2017-2024</i>	46
Figura 9 <i>Tendencia Cicatrices de Quema 2017-2024</i>	47
Figura 10 <i>Tendencia Cabezas de Ganado 2017-2024</i>	48
Figura 11 <i>Flujo ETL</i>	49
Figura 12 <i>Plataforma de Google Earth Engine</i>	53
Figura 13 <i>Clasificación de Coberturas con Dynamic World</i>	54
Figura 14 <i>Dataset PNN Tinigua_Macarena</i>	54
Figura 15 <i>Distribución de las Variables PNN Chibiriquete</i>	56
Figura 16 <i>Boxplots Variables PNN Chibiriquete</i>	58
Figura 17 <i>Distribución de las Variables PNN Tinigua-Macarena</i>	60
Figura 18 <i>Boxplots Variables PNN Tinigua-Macarena</i>	62
Figura 19 <i>Distribuciones Suavizadas</i>	65
Figura 20 <i>Resultado Regresión Lineal Múltiple</i>	67
Figura 21 <i>Correlación Variables para PNN Chibiriquete</i>	69
Figura 22 <i>Resultado Modelo GAM</i>	71

Figura 23 <i>Validación Cruzada del Modelo GAM</i>	73
Figura 24 <i>Validación Cruzada con Ajuste de Parámetros del Modelo GAM</i>	75
Figura 25 <i>Modelo de Regresión Lineal Múltiple para PNN Tinigua-Macarena</i>	77
Figura 26 <i>Correlación Variables para PNN Tinigua-Macarena</i>	81
Figura 27 <i>Modelo Ridge para PNN Tinigua-Macarena</i>	83
Figura 28 <i>Validación Cruzada para el Modelo Ridge</i>	84

Introducción

La deforestación en la Amazonía colombiana constituye un desafío ambiental cuyos efectos sobre la biodiversidad, los ciclos hidrológicos y el equilibrio climático global son de gran importancia. Este fenómeno, con un alto crecimiento en las últimas décadas, está afectando de manera notoria a los parques nacionales de la región, en particular al Parque Nacional Natural Serranía del Chibiriquete, el Parque Nacional Natural Tinigua y la Sierra de la Macarena (A. León-Parra et al., 2023). A pesar de los esfuerzos de conservación impulsados por las autoridades y entidades no gubernamentales, la falta de datos precisos y actualizados sobre el fenómeno impide una respuesta oportuna y efectiva (A. León-Parra et al., 2023).

Este proyecto se propone analizar los factores de riesgo de deforestación en estas áreas protegidas mediante la utilización de datos de puntos de calor, cicatrices de quema, explotación ganadera y praderización, con el propósito de facilitar la identificación de zonas de alto riesgo y promover estrategias de conservación y manejo proactivo.

Los resultados de este análisis pretenden mostrar como el uso de nuevas tecnologías y herramientas pueden contribuir al monitoreo actualizado de las zonas protegidas, y que la creación de indicadores de riesgo, fundamentados en evidencia técnica, pueden servir para guiar las políticas y esfuerzos de conservación de manera más efectiva.

Planteamiento del Problema

La acelerada deforestación en la Amazonía colombiana compromete de manera grave la biodiversidad única de la región y contribuye significativamente al cambio climático, afectando tanto el equilibrio ecológico como el socioeconómico. Los parques nacionales de Chibiriquete, Tinigua y la Sierra de la Macarena enfrentan una presión creciente debido a la expansión de actividades humanas, como la ganadería y los cultivos ilícitos (Marina et al., 2017). Además de los puntos de calor, cicatrices de quema y degradación del bosque, los investigadores del fenómeno de la deforestación del Amazonas colombiano han identificado otros factores importantes que influyen, tales como las carreteras terciarias, las trochas, la praderización de zonas del bosque para hacer ganadería extensiva, la plantación de cultivos lícitos e ilícitos, y la apropiación de tierras (Inés González & Gómez, 2020; Marina et al., 2017). Lamentablemente, no hay datos suficientes acerca de estos factores; los datos disponibles son fragmentarios y es difícil mantenerlos actualizados. El Instituto SINCHI provee cierta información de manera actual, pero otros datos que son determinantes para evaluar el fenómeno, como la degradación del bosque, la praderización y la pérdida de cobertura, suelen tener un rezago de entre uno y dos años debido a la metodología empleada para medirlos, entre otros factores.

Este proyecto tiene como objetivo superar esta barrera mediante el uso de tecnologías modernas de monitoreo ambiental y análisis de datos, con el fin de desarrollar indicadores que permitan la identificación de las áreas más vulnerables en tiempo casi real.

La problemática se agrava debido a la falta de un monitoreo constante que permita una respuesta rápida frente a la acelerada degradación de estos ecosistemas. Los procesos actuales de deforestación se encuentran vinculados no solo a actividades ganaderas, sino también a factores como el desarrollo de vías ilegales, el aumento de los asentamientos humanos y la expansión de

la frontera agrícola. Estos factores, en conjunto, generan un panorama complejo donde cada elemento actúa como un catalizador de la pérdida de cobertura forestal, dificultando la implementación de políticas de conservación efectivas. que permita una respuesta rápida frente a la acelerada degradación de estos ecosistemas.

La deforestación también tiene un impacto socioeconómico profundo en las comunidades locales, que dependen de los recursos naturales de la Amazonía para su subsistencia. La pérdida de servicios ecosistémicos, tales como la regulación del ciclo del agua, la provisión de alimentos y la captura de carbono, afecta directamente la calidad de vida de estas comunidades, aumentando su vulnerabilidad y limitando sus oportunidades de desarrollo sostenible (Marina et al., 2017). Ante esta situación, se hace imprescindible contar con herramientas de análisis que permitan entender la dinámica de la deforestación de manera integral y actuar con la urgencia que requiere la protección de estos territorios.

Justificación

La Amazonía colombiana, conocida como el "pulmón del mundo", enfrenta amenazas graves derivadas de la deforestación creciente, con implicaciones directas tanto en los ecosistemas locales como en el clima global. Esta situación no solo afecta la biodiversidad y los recursos naturales, sino que también tiene un impacto profundo en la estabilidad ecológica y el equilibrio climático del planeta. Las tecnologías actuales de observación de la Tierra, como Google Earth Engine y Dynamic World, ofrecen la posibilidad de monitorear los ecosistemas y generar datos valiosos prácticamente en tiempo real, contribuyendo significativamente al entendimiento de los cambios en la cobertura terrestre. Estas herramientas permiten obtener una visión más precisa y actualizada del entorno, lo cual es fundamental para la conservación. Estas herramientas permiten obtener información detallada sobre la degradación del bosque, como la praderización, lo cual representa un avance significativo en comparación con las metodologías tradicionales.

El problema radica en que los datos proporcionados por instituciones gubernamentales y no gubernamentales no están actualizados. Estos datos presentan un rezago de hasta dos años, lo cual dificulta la implementación de respuestas efectivas y oportunas. Este proyecto se justifica precisamente porque busca ofrecer alternativas a esta limitación utilizando herramientas tecnológicas avanzadas que permitan la obtención de datos casi en tiempo real. La integración de datos provenientes de puntos de calor, cicatrices de quema y explotación ganadera, junto con la información de degradación del bosque obtenida mediante Google Earth Engine, tiene el potencial de ofrecer una visión más completa y actualizada del fenómeno de la deforestación en los parques nacionales Chibiriquete, Tinigua y Sierra de la Macarena.

El uso de nuevas herramientas y tecnologías permitirá desarrollar indicadores de riesgo oportunos y establecer sistemas de alerta temprana para la identificación de áreas críticas que requieran atención inmediata. Estos indicadores serán útiles para las comunidades locales, ya que les proporcionarán información clara sobre las zonas de riesgo, permitiéndoles adaptar sus actividades y prácticas de manejo para minimizar el impacto ambiental. Además, facilitarán la toma de decisiones por parte de los actores responsables de la conservación, como agencias gubernamentales y ONG, al orientar las intervenciones hacia las áreas que más lo necesitan. Esto contribuirá a la conservación del medio ambiente y fortalecerá las capacidades locales de resiliencia y adaptación frente a los cambios ambientales. Además, se busca aportar una metodología replicable para el monitoreo de otras áreas protegidas, promoviendo la gestión sostenible de los recursos naturales.

El enfoque de este estudio se justifica no solo por la importancia ecológica de la Amazonía, sino también por su relevancia socioeconómica. Las comunidades locales dependen de estos ecosistemas para su subsistencia y bienestar, y la deforestación amenaza directamente su calidad de vida. Al mejorar la capacidad para identificar áreas vulnerables y actuar de manera oportuna, este estudio también pretende fortalecer la resiliencia de estas comunidades y promover prácticas de manejo sostenible que aseguren tanto la protección del medio ambiente como el bienestar socioeconómico de los habitantes de la región. Algunas de las prácticas específicas incluyen la agroforestería, que combina cultivos con árboles para preservar la biodiversidad y mejorar la fertilidad del suelo, el uso de pastoreo rotacional para reducir la presión sobre el terreno y evitar la sobreexplotación, y la promoción de cultivos sostenibles que minimicen el impacto ambiental y fomenten la regeneración de los ecosistemas.

Por último, la aplicación de enfoques tecnológicos como el análisis geoespacial, el uso de sensores satelitales y modelos estadísticos permitirá desarrollar una metodología replicable, no solo en otras áreas de la Amazonía, sino también en diferentes regiones del mundo que enfrentan problemas similares. Este carácter replicable y escalable del proyecto incrementa su relevancia y lo convierte en un aporte valioso para la conservación a nivel global.

El enfoque de este estudio se justifica no solo por la importancia ecológica de la Amazonía, sino también por su relevancia socioeconómica. Las comunidades locales dependen de estos ecosistemas para su subsistencia y bienestar, y la deforestación amenaza directamente su calidad de vida. La implementación de políticas y estrategias basadas en la evidencia permitirá una mayor eficiencia en los esfuerzos de conservación, optimizando el uso de recursos y fomentando la colaboración entre diferentes actores involucrados en la protección de la Amazonía.

Objetivos

Objetivo General

Analizar espacial y temporalmente los factores de riesgo de deforestación en los parques nacionales Chibiriquete, Tinigua y Sierra de la Macarena, utilizando datos actualizados de puntos de calor, cicatrices de quema y explotación ganadera, con el propósito de evaluar su suficiencia para identificar áreas prioritarias y apoyar la toma de decisiones en conservación de manera más oportuna.

Objetivos Específicos

Describir de forma espacial y temporal los factores de riesgo de deforestación alrededor de los parques nacionales naturales Chibiriquete, Tinigua y Sierra de la Macarena.

Revisar los datos actualizados de puntos de calor, cicatrices de quema y explotación ganadera, asegurando su calidad y compatibilidad espacial y temporal.

Identificar las áreas de praderización en los parques nacionales chibiriquete, tinigua y sierra de la macarena.

Realizar un análisis exploratorio de datos para comprender la distribución y variación de los factores de riesgo.

Definir un indicador que refleje el riesgo de deforestación (IRD) en diferentes áreas y momentos, basados en los datos actualizados.

Marco Conceptual

La deforestación se refiere a la eliminación o reducción significativa de áreas forestales, a menudo para dar lugar a otros usos de la tierra como la agricultura, la ganadería, la minería o la urbanización. Este proceso puede ser resultado de actividades humanas o de fenómenos naturales.

Causas de la Deforestación

Existen diversas razones que conducen a la deforestación:

Expansión agrícola: La necesidad de crear nuevas tierras de cultivo para satisfacer la demanda de alimentos y otros productos agrícolas.

Ganadería: La expansión de pastos para la cría de ganado, especialmente en regiones tropicales.

Minería: La extracción de recursos minerales que requiere la eliminación de la vegetación superficial.

Urbanización: El crecimiento de áreas urbanas que demanda espacio para viviendas, infraestructura y servicios.

Tala de árboles: La explotación maderera, tanto legal como ilegal, para obtener madera y otros productos forestales.

Consecuencia de la Deforestación

La deforestación tiene múltiples impactos negativos:

Pérdida de biodiversidad: Los bosques albergan la mayoría de las especies terrestres; su destrucción implica la pérdida de hábitats y, potencialmente, la extinción de especies.

Cambio climático: Los árboles actúan como sumideros de carbono, absorbiendo CO₂. Al ser talados, liberan el carbono almacenado, contribuyendo al aumento de gases de efecto invernadero y al calentamiento global

Erosión del suelo: La ausencia de vegetación aumenta la susceptibilidad del suelo a la erosión, lo que puede degradar la tierra y causar deslizamientos.

Alteración del ciclo del agua: Los árboles desempeñan un papel crucial en la regulación del ciclo hidrológico; su eliminación puede modificar patrones de lluvia y disponibilidad de agua.

Métodos y Variables para Medir la Deforestación en Colombia

El Instituto SINCHI, una entidad colombiana dedicada a la investigación y gestión de información sobre la Amazonia, desempeña un papel fundamental en el estudio y monitoreo de este ecosistema único. A través de su plataforma de ArcGIS Online, el Instituto SINCHI proporciona una amplia gama de recursos cartográficos y datos geoespaciales que son cruciales para la comprensión de los procesos de cambio en la Amazonia colombiana.

Uno de los aspectos fundamentales de los datos proporcionados por el Instituto SINCHI es su utilidad para identificar y analizar patrones de cambio en la cobertura terrestre, como la degradación del bosque, la praderización, y la expansión de zonas de pastizales. Estos cambios a menudo están relacionados con procesos humanos como la ganadería extensiva y la agricultura, que a su vez son factores de riesgo para la deforestación en la Amazonia. A través del uso de plataformas como ArcGIS, el Instituto SINCHI facilita el acceso a estos datos, promoviendo una mayor transparencia y colaboración entre diferentes actores, incluidos investigadores, organizaciones no gubernamentales y las autoridades locales.

Se relación a continuación las variables de la deforestación que serán empleadas en el desarrollo del presente trabajo:

Puntos de Calor

Los puntos de calor son áreas identificadas mediante el análisis de datos térmicos provenientes de sensores satelitales como MODIS (Moderate Resolution Imaging Spectroradiometer) y VIIRS (Visible Infrared Imaging Radiometer Suite). Estos sensores capturan imágenes en bandas espectrales que permiten detectar temperaturas anormalmente altas en la superficie terrestre.

Los puntos de calor se asocian comúnmente con incendios o actividades humanas que generan elevadas temperaturas, y son utilizados como un indicador temprano de deforestación o quema en curso.

La metodología para la identificación de puntos de calor incluye la selección de imágenes satelitales adecuadas, el preprocesamiento de las imágenes (incluyendo corrección radiométrica y geométrica), y la aplicación de umbrales de temperatura que permiten distinguir áreas normales de aquellas con temperaturas elevadas.

Cicatrices de Quema

Las cicatrices de quema son áreas afectadas por incendios forestales y se identifican utilizando técnicas de teledetección, generalmente a partir de imágenes de satélites como Landsat 8 o Sentinel-2. Estas imágenes permiten calcular índices espectrales, como el Índice de Área Quemada (BAI) y el Índice de Vegetación de Diferencia Normalizada (NDVI), que resaltan las áreas que han sufrido daños por el fuego. El proceso para generar mapas de cicatrices de quema incluye la selección de imágenes con poca cobertura de nubes, el preprocesamiento para mejorar

la calidad de los datos espectrales, y la clasificación de las áreas quemadas mediante algoritmos supervisados o no supervisados.

Pérdida de Cobertura Forestal

Se refiere a la disminución de la superficie cubierta por bosques en un área determinada. Este proceso puede ser consecuencia de diversas actividades humanas, como la tala de árboles, la expansión agrícola, la minería y la urbanización. La pérdida de cobertura forestal es un indicador directo de la deforestación y se evalúa mediante imágenes satelitales.

Praderización

La praderización se refiere a la expansión de áreas de pastos, generalmente a expensas de los bosques naturales, con el fin de establecer sistemas ganaderos extensivos. Este fenómeno es una de las principales causas de la deforestación en la Amazonia colombiana. Para identificar y cuantificar las áreas afectadas por la praderización, se utilizan imágenes satelitales de alta resolución, como las de Landsat y Sentinel. Estas imágenes se procesan para realizar clasificaciones de cobertura terrestre y se comparan de forma multitemporal para detectar cambios. El análisis multitemporal permite calcular la tasa media anual de praderización (TMAP), la cual indica la velocidad de expansión de las áreas de pastos en un período determinado.

Metodologías de Clasificación de Cobertura Terrestre

Históricamente, la clasificación de la cobertura terrestre se ha realizado utilizando metodologías como Corine Land Cover (CLC), reconocida por su robustez y estandarización a nivel internacional. Sin embargo, la aparición de nuevas herramientas como Dynamic World y Google Earth Engine (GEE) ha permitido un enfoque más dinámico y flexible para el análisis de coberturas terrestres.

Cada una de estas herramientas y metodologías tiene características específicas, ventajas y desventajas que las hacen adecuadas para diferentes tipos de análisis:

Dynamic World: Es un conjunto de datos globales que proporciona información sobre el uso y la cobertura de la tierra en tiempo casi real, con una resolución de 10 metros. Este recurso ha sido desarrollado en colaboración entre Google y el Instituto de Recursos Mundiales (World Resources Institute, WRI).

Dynamic World utiliza imágenes satelitales del programa Sentinel-2 de la Agencia Espacial Europea. Mediante técnicas de aprendizaje profundo (deep learning), clasifica cada píxel de las imágenes en una de nueve categorías de cobertura terrestre: agua, árboles, pastos, cultivos, vegetación inundada, áreas construidas, suelo desnudo, nieve y hielo, y arbustos.

El proceso incluye:

Adquisición de imágenes: Se utilizan imágenes de Sentinel-2 con una cobertura de nubes inferior al 35%.

Preprocesamiento: Se aplican correcciones radiométricas y geométricas para mejorar la calidad de las imágenes.

Clasificación: Un modelo de aprendizaje profundo asigna probabilidades a cada píxel para las nueve clases de cobertura.

Posprocesamiento: Se generan mapas de cobertura aplicando umbrales de probabilidad y técnicas de filtrado espacial.

Google Earth Engine (GEE): Es una plataforma de análisis geoespacial en la nube que permite acceder a un extenso catálogo de datos satelitales y realizar análisis personalizados. Su ventaja principal es la flexibilidad que ofrece al permitir el desarrollo de algoritmos de clasificación adaptados a las necesidades del usuario, además de la capacidad de procesamiento a

gran escala. Sin embargo, requiere conocimientos técnicos en programación, y la calidad de los resultados depende de los algoritmos utilizados y de la calidad de las imágenes disponibles.

Corine Land Cover (CLC): La metodología CLC es un sistema estandarizado desarrollado por la Agencia Europea de Medio Ambiente, adaptado en Colombia para mapear la cobertura terrestre. Su principal ventaja es la estandarización y comparabilidad internacional, lo cual facilita estudios comparativos a nivel regional. Además, proporciona datos históricos desde 1990, permitiendo análisis multitemporales. No obstante, presenta una resolución espacial limitada (100 metros), lo que puede no ser adecuado para estudios que requieren un nivel de detalle más alto, y su actualización es menos frecuente en comparación con otras metodologías.

Tabla 1

Comparación Tecnologías de Monitoreo del Terreno

Metodología	Ventajas	Desventajas
Dynamic World	Actualización continua, alta resolución espacial (10 m), consistencia global. Algoritmo de deep learning para determinar el tipo de cobertura de suelo de un terreno.	Acceso limitado a datos históricos (datos desde 2015), afectado por la cobertura de nubes.
Google Earth Engine	Flexibilidad para desarrollar algoritmos personalizados, capacidad de procesamiento masivo, acceso a múltiples datasets.	Requiere conocimientos técnicos en programación, variabilidad en la calidad de datos.
Corine Land Cover	Estandarización y comparabilidad internacional, datos históricos desde 1990.	Resolución espacial limitada (100 m), actualización infrecuente, aplicabilidad limitada fuera de Europa.

Nota. Ventajas y desventajas de las tecnologías para el monitoreo de la deforestación.

Pertinencia Actual de los Datos

En la actualidad los datos suministrados por el instituto SINCHI que presentan una actualización casi en tiempo real son los correspondientes a puntos de calor y cicatrices de quema.

En contraste, los datos relacionados con coberturas de suelo, pérdida de bosque, degradación de bosques y praderización presentan, dada la metodología de medición un rezago de entre uno y dos años.

Cabezas de Ganado

De acuerdo con la evidencia la explotación ganadera es uno de los factores conductores de la deforestación en diferentes regiones del país, principalmente en la amazonía colombiana.

El Sistema de Vigilancia Epidemiológica en el sector pecuario implementado en Colombia es una estrategia esencial para la detección de enfermedades y llevar a cabo programas de prevención, control o erradicación de estas.

Este sistema permite documentar rigurosamente las cabezas de ganado en las diferentes regiones del país y, por lo tanto, los datos relacionados se consideran actualizados y aptos para el desarrollo de análisis con respecto de los fenómenos de deforestación en la amazonía.

En resumen, el marco conceptual de este estudio se basa en el uso de técnicas de teledetección y el análisis de datos geospaciales proporcionados por plataformas tradicionales y tecnológicas de vanguardia Google Earth Engine y Dynamic World sumado a datos socioeconómicos como es la explotación ganadera. Estos recursos permiten una mejor comprensión de los factores de riesgo que impulsan la deforestación y la degradación del bosque en la Amazonia colombiana, y constituyen herramientas fundamentales para la toma de decisiones en la conservación de este ecosistema.

Marco Teórico

El presente trabajo se enfoca en la aplicación de técnicas de ciencia de datos para la modelización y análisis de los factores de riesgo que inciden en la deforestación y la degradación de los bosques en la Amazonía colombiana. En esta sección se presenta un marco teórico que abarca las metodologías y fundamentos estadísticos esenciales utilizados en este estudio.

Ciencia de Datos en el Contexto de la Deforestación

La ciencia de datos representa un enfoque interdisciplinario que permite extraer información significativa y generar conocimiento a partir de datos complejos. En el contexto del estudio de la deforestación, la ciencia de datos proporciona herramientas sofisticadas para el análisis de grandes volúmenes de datos geoespaciales, posibilitando la identificación de patrones temporales y espaciales relacionados con la pérdida de cobertura forestal. Esta aproximación se implementa a través de herramientas como Google Earth Engine y Dynamic World, que permiten el acceso y procesamiento eficiente de datos satelitales, facilitando la construcción de modelos precisos que describen las dinámicas forestales. Históricamente, la metodología Corine Land Cover (CLC) ha sido ampliamente utilizada debido a su robustez y estandarización a nivel internacional. Sin embargo, la aparición de nuevas herramientas como Dynamic World y Google Earth Engine ha permitido un enfoque más dinámico y flexible para el análisis de coberturas terrestres, proporcionando datos más actualizados y con mayor resolución.

Estadística Descriptiva

La estadística descriptiva es una parte fundamental del análisis exploratorio de datos, que permite resumir y describir las características principales de los datos recolectados. En este trabajo, se aplicarán métricas como la media, la mediana, los cuartiles, la desviación estándar y la distribución de frecuencias, con el propósito de proporcionar una visión general de las

variables estudiadas, tales como puntos de calor, cicatrices de quema y áreas praderizadas.

Asimismo, se emplearán representaciones gráficas, como histogramas y diagramas de caja (boxplots), para identificar tendencias, valores atípicos y posibles sesgos presentes en los datos.

Análisis Temporal

El análisis temporal es crucial para comprender la evolución de los fenómenos asociados a la deforestación a lo largo del tiempo. En este estudio, se llevará a cabo un análisis de la evolución de las variables de interés (por ejemplo, puntos de calor y cicatrices de quema) durante los diferentes cuatrimestres del período de estudio. Este análisis permitirá identificar patrones estacionales, ciclos de actividad y eventos anómalos, contribuyendo a una mejor comprensión de las fluctuaciones en la deforestación en los parques nacionales objeto de estudio.

Modelos de Relación entre Variables - Regresión Lineal Múltiple

La regresión lineal múltiple es una técnica estadística que permite evaluar la relación entre una variable dependiente y múltiples variables independientes. En este contexto, la variable dependiente es la degradación del bosque, representada por la praderización, mientras que las variables independientes incluyen los puntos de calor, las cicatrices de quema y las cabezas de ganado. La regresión lineal múltiple proporciona una estimación cuantitativa de la influencia de cada una de estas variables sobre la degradación del bosque, facilitando una visión integral de los factores subyacentes que contribuyen a la deforestación.

Modelos Generalizados Aditivos (GAM)

Los modelos generalizados aditivos (GAM) son herramientas que permiten capturar relaciones no lineales entre la variable dependiente y las variables independientes. Debido a que las relaciones entre los factores de riesgo y la degradación del bosque no siempre son lineales, los GAM resultan adecuados para modelar estas complejas interacciones de manera flexible. En

este estudio, se utilizarán GAM para explorar cómo cada factor contribuye a la degradación del bosque y para desarrollar un indicador de riesgo de deforestación, el cual permitirá evaluar el estado de cada municipio y cuatrimestre dentro del parque nacional.

Análisis de Residuos y Validación de Modelos

La validación de modelos y el análisis de residuos son pasos esenciales para garantizar la calidad y la precisión de los modelos utilizados. En este trabajo, se realizará un análisis detallado de los residuos generados por los modelos de regresión lineal y GAM, con el fin de evaluar el ajuste del modelo y la distribución de los errores. Este análisis contribuirá a la identificación de problemas como la heterocedasticidad y la autocorrelación, y permitirá realizar ajustes en los modelos para mejorar su capacidad predictiva.

Pertinencia Actual de los Datos

En la actualidad, los datos suministrados por el Instituto SINCHI que presentan una actualización casi en tiempo real son los correspondientes a puntos de calor y cicatrices de quema. En contraste, los datos relacionados con coberturas de suelo, pérdida de bosque, degradación de bosques y praderización presentan, dada la metodología de medición, un rezago de entre uno y dos años. Esta situación evidencia la importancia de la integración de tecnologías como Google Earth Engine y Dynamic World, que ofrecen datos más actualizados y permiten complementar la información proporcionada por el Instituto SINCHI

El marco teórico de este estudio se fundamenta en la integración de conceptos y metodologías avanzadas de estadística y ciencia de datos para evaluar de manera exhaustiva los factores de riesgo asociados a la deforestación. La combinación de análisis descriptivos, análisis temporales y modelos de regresión permite una comprensión profunda de las dinámicas que

impulsan la degradación del bosque, contribuyendo al desarrollo de estrategias informadas para la conservación y el manejo sostenible de los recursos naturales en la Amazonía colombiana.

Metodología

Este estudio emplea un enfoque cuantitativo, basado en el análisis espaciotemporal de datos actualizados. Se eligió este enfoque porque permite una evaluación objetiva y medible de los factores de riesgo de deforestación, facilitando la identificación de patrones y relaciones que no serían evidentes con enfoques cualitativos. Además, el enfoque cuantitativo es ideal para analizar grandes volúmenes de datos geospaciales y generar resultados replicables y estadísticamente significativos.

El análisis se realizará de manera diferenciada para el Parque Nacional Serranía del Chibiriquete y, por separado, de manera conjunta, para los Parques Tinigua y la Sierra de la Macarena, con el propósito de evaluar con mayor precisión los factores de riesgo y sus efectos.

Las variables para considerar para el estudio son los puntos de calor, las cicatrices de quema, las cabezas de ganado por municipio y, como variable de estudio o dependiente se trabajará con la praderización, como variable de medición de la degradación del bosque.

Pasos Metodológicos

De acuerdo con lo anterior, los pasos metodológicos para el desarrollo del presente trabajo son los siguientes:

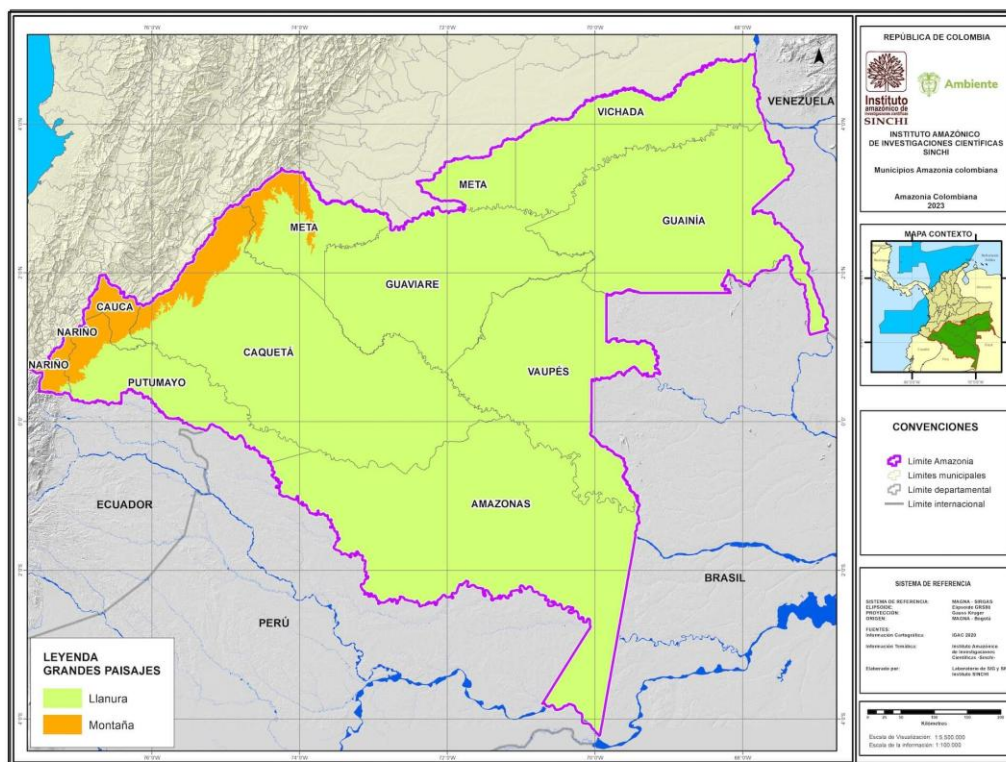
1. Definición de las áreas de análisis.
2. Procesamiento, exploración y análisis de datos empleando técnicas de ciencia de datos.
3. Formulación de indicadores de riesgo de deforestación para las áreas definidas.

Definición de las Áreas de Análisis

El área de estudio está centrada en la amazonía colombiana.

Figura 1

Delimitación de la Amazonía Colombiana



Nota. Visualización del área que ocupa la amazonía colombiana y los departamentos que la componen. Fuente (Sistema de información territorial amazonía colombiana, 2022).

En esta zona se encuentran numerosos parques naturales nacionales, resguardos indígenas y áreas protegidas. Los parques sobre los cuales se centra el análisis son los parques naturales nacionales serranía del Chibiriquete, Tinigua y sierra de la Macarena.

Los departamentos en los cuales se ubican los parques relacionados son Caquetá, Meta y Guaviare.

A su vez, los municipios relacionados con estos departamentos y parques nacionales son:

- **Caquetá:** Cartagena del Chairá, San Vicente del Caguán y Solano.

- Guaviare: Calamar, Miraflores y San José del Guaviare.
- Meta: La Macarena, Mesetas, San Juan de Arama, Puerto Rico, Uribe y

Vistahermosa.

Figura 2

Parques Nacionales Naturales



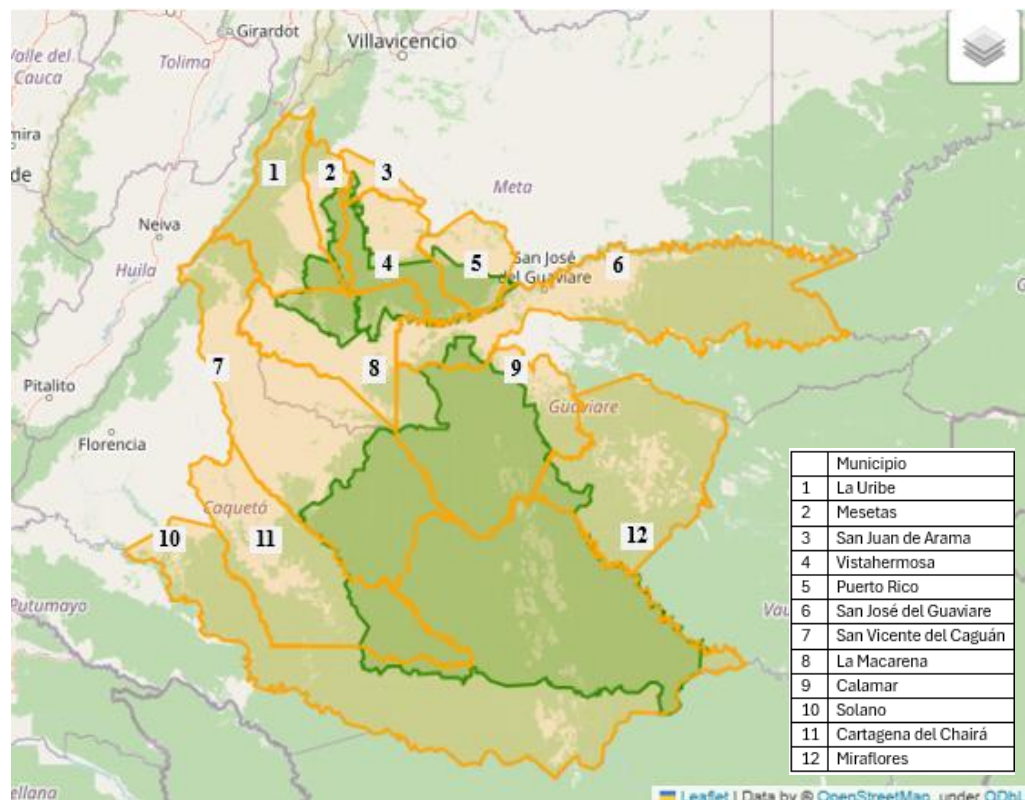
Nota. Visualización de la ubicación de los parques nacionales de interés y los departamentos donde se ubican.

Los municipios se convierten en el eje central de configuración de los datos para los análisis posteriores. Los puntos de calor y los datos de cabezas de ganado están referenciados por municipio. Se propone llevar los datos de cicatrices de quema y praderización con marco de

referencia igualmente por municipio para poder configurar un conjunto de datos congruente para el estudio.

Figura 3

Municipios en PNN Chibiriquete, Tinigua y la Macarena



Nota. Se visualiza la ubicación de los diferentes municipios alrededor de los parques nacionales naturales del estudio.

Finalmente, se establecen los municipios para cada parque nacional, que sirven como foco de datos para el estudio de la deforestación en dichos parques, de la siguiente manera:

- PNN Chibiriquete: San Vicente del Caguán, Cartagena del Chairá, Solano, Calamar, Miraflores y San José del Guaviare.

- PNN Tinigua-Macarena: La Uribe, La Macarena, Mesetas, San Juan de Arama, Vistahermosa y Puerto Rico.

Procesamiento, Exploración y Análisis de Datos

En esta etapa se emplean las herramientas específicas de software como Python y R para llevar a cabo la recolección y procesamiento de datos. Posteriormente, las técnicas estadísticas como la estadística descriptiva permiten hacer la exploración de los datos para entender su comportamiento y, finalmente, el estudio de la relación entre las variables por medio modelos de regresión lineal, regularización de coeficientes por medio del método Ridge y Lasso y exploración de modelos de regresión no lineales, permite una comprensión más amplia de las dinámicas que impulsan la deforestación en estas áreas protegidas.

Las herramientas como R y Python son adecuadas para el análisis estadístico y el procesamiento de datos geoespaciales. R es especialmente útil por sus poderosas bibliotecas para análisis estadístico y visualización de datos, mientras que Python ofrece gran flexibilidad y capacidades avanzadas en procesamiento de datos geoespaciales mediante bibliotecas como Pandas y GeoPandas. Esta combinación permite un análisis más completo y eficiente.

Los pasos específicos en esta etapa son:

1. Recolección y preparación de datos.
2. Análisis espaciotemporal.
3. Análisis exploratorio de datos.
4. Desarrollo y análisis de modelos para estudio de relación de las variables.

A continuación, se describen de los componentes de dichos pasos.

Recolección y Preparación de Datos

Se describen a continuación las fuentes y características de los datos propuestos para la realización del análisis.

Datos de puntos de calor:

- Fuente: Datos abiertos proporcionados por el SIAT-AC. Estos datos a su vez se obtienen por medio de Sensores satelitales MODIS y VIIRS.
- Frecuencia: Diaria. Datos disponibles desde el 2002.

Datos de cicatrices de quema:

- Fuente: Datos abiertos proporcionados por el SIAT-AC. Estos datos a su vez se obtienen a partir de imágenes satelitales de Sentinel-2 y Landsat 8.
- Frecuencia: Datos mensuales procesados mediante índices espectrales (NBR).

Datos disponibles desde el 2017.

Datos de explotación ganadera:

- Fuente: Registros oficiales de cabezas de ganado proporcionados por el ICA (Instituto Colombiano Agropecuario). Datos de programas de vacunación para ganado bovino (DANE).
- Frecuencia: Anual. Datos disponibles desde el 2012.

Datos de praderización en PNN:

- Fuente: Datos de clasificación de tipo de cobertura de suelo y cálculo de áreas de praderización. Se obtienen por medio del algoritmo de clasificación de Dynamic World aplicado en la plataforma de Google Earth Engine para las zonas de interés.
- Frecuencia: Semanal. Datos disponibles desde el 2015.

Análisis Espaciotemporal

En esta etapa se realiza el mapeo de los indicadores de riesgo a lo largo del tiempo, generando mapas y gráficas de tendencia que muestren su comportamiento.

Análisis Exploratorio de Datos

En esta etapa se determina la representatividad y suficiencia de los datos actualizados para el análisis del riesgo de deforestación.

Igualmente, se elaboran mapas de distribución espacial de puntos de calor y cicatrices de quema. Estos mapas serán fundamentales para identificar las zonas más afectadas y para comunicar visualmente los resultados a actores clave.

Desarrollo y Análisis de Modelos

En esta etapa se desarrollan los modelos que permiten estudiar y establecer la relación entre variables de estudio. También, se realiza el análisis de dichos modelos para asegurar la robustez de los resultados obtenidos.

Formulación de Indicadores de Riesgo

En esta última etapa de la metodología planteada se proponen y desarrollan indicadores de riesgo de la deforestación IRD, que permitan medir y comunicar de manera clara y objetiva el riesgo de deforestación para las áreas de estudio.

Descripción Espacial y Temporal de los Factores de Riesgo

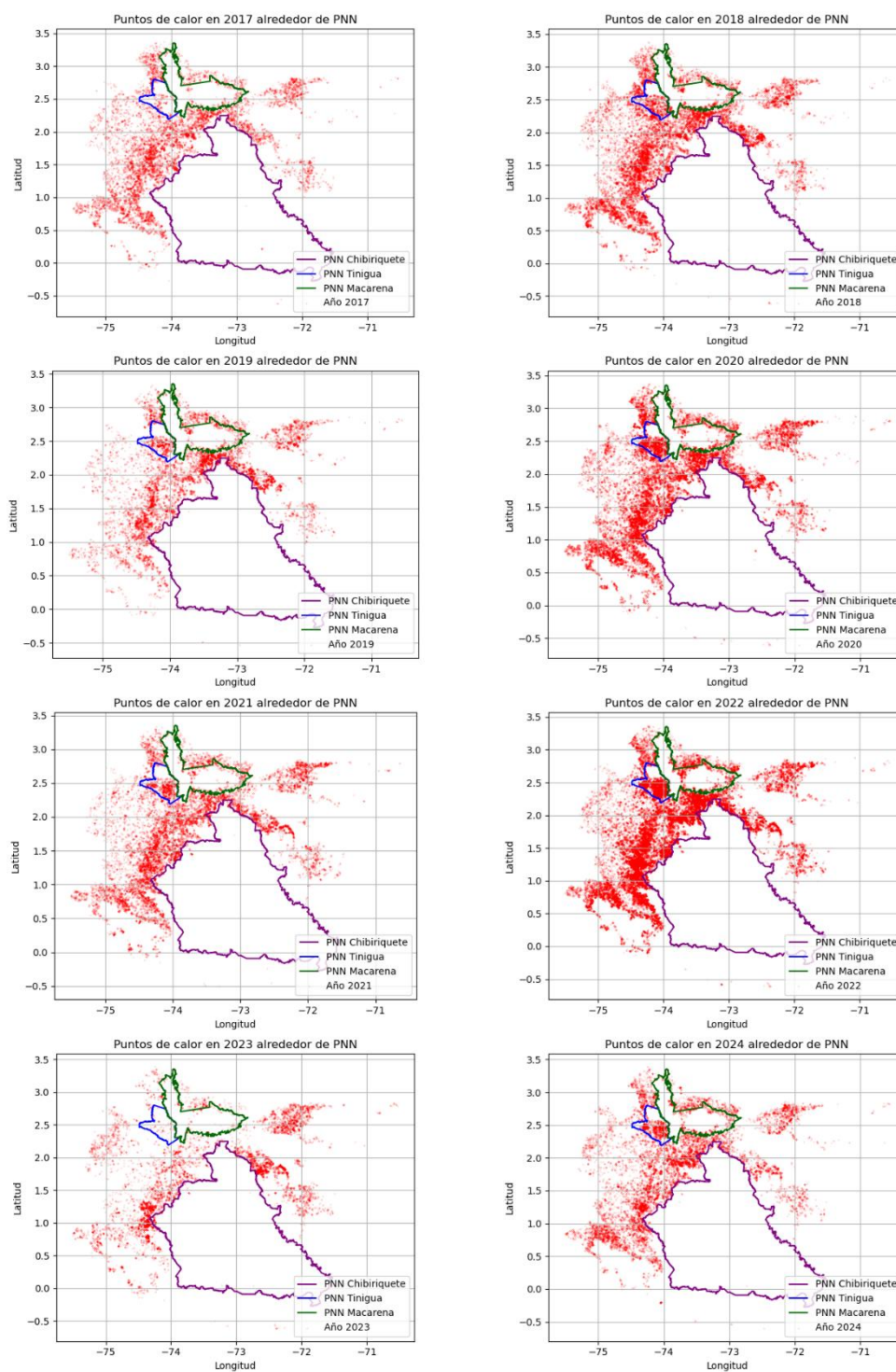
La descripción espaciotemporal de los factores de riesgo de deforestación resulta esencial para comprender cómo los patrones de actividades humanas afectan la integridad de áreas protegidas, como los Parques Nacionales Chibiriquete y Tinigua-Macarena. Este tipo de exploración facilita la identificación de zonas particularmente vulnerables y de los momentos en que se intensifican los riesgos, proporcionando información valiosa para diseñar estrategias de mitigación y conservación ajustadas a la dinámica del entorno amazónico colombiano. A través del estudio de los factores de riesgo, como los puntos de calor, las cicatrices de quema y la ganadería extensiva, este análisis ayuda a determinar los impulsores más críticos de la degradación forestal para así poder priorizar esfuerzos de conservación.

Puntos de Calor 2017-2024

Las visualizaciones de puntos de calor alrededor de los Parques Nacionales Chibiriquete y Tinigua-Macarena durante el periodo de 2017 a 2024 muestran una concentración sostenida de estos eventos, con incrementos más pronunciados en los años 2020 y 2022. Esta tendencia sugiere una presión continua sobre las áreas forestales. Los boxplots que ilustran las distancias de los puntos de calor a los parques para cada municipio evidencian una estabilidad en la mediana de las distancias con una significativa dispersión de valores. Esto indica la ocurrencia de eventos esporádicos que están más cerca o más lejos del parque de lo esperado. Municipios como Solano, San José del Guaviare y La Macarena presentan un mayor número de valores atípicos, lo cual podría estar relacionado con picos en la actividad humana durante ciertos periodos.

Figura 4

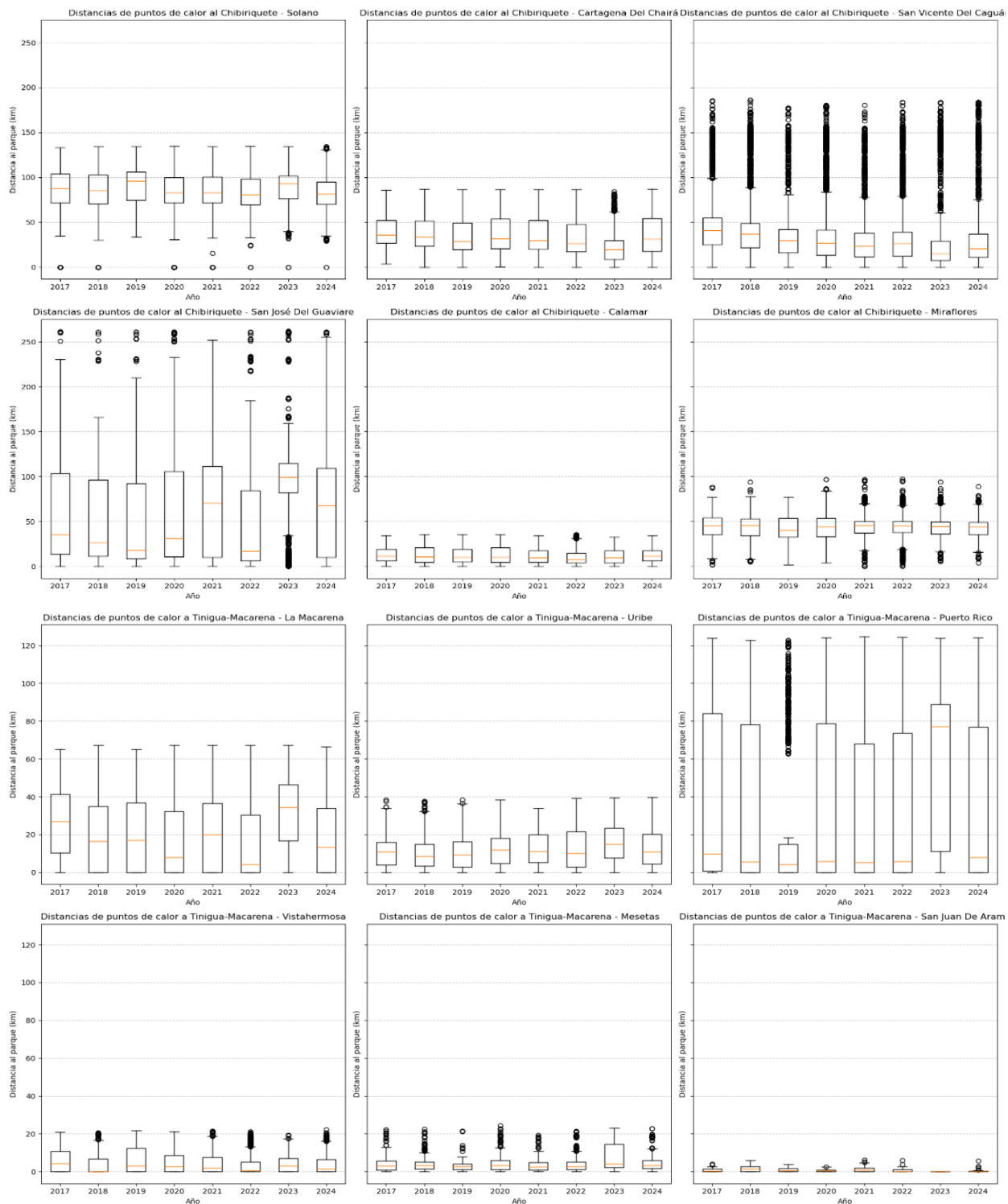
Comportamiento Temporal Puntos de Calor en PNN



Nota. Comportamiento puntos de calor alrededor de PNN.

Figura 5

Boxplots Distancia Puntos de Calor a PNN



Nota. Visualización acercamiento de puntos de calor sobre PNN.

Cicatrices de Quema 2017-2024

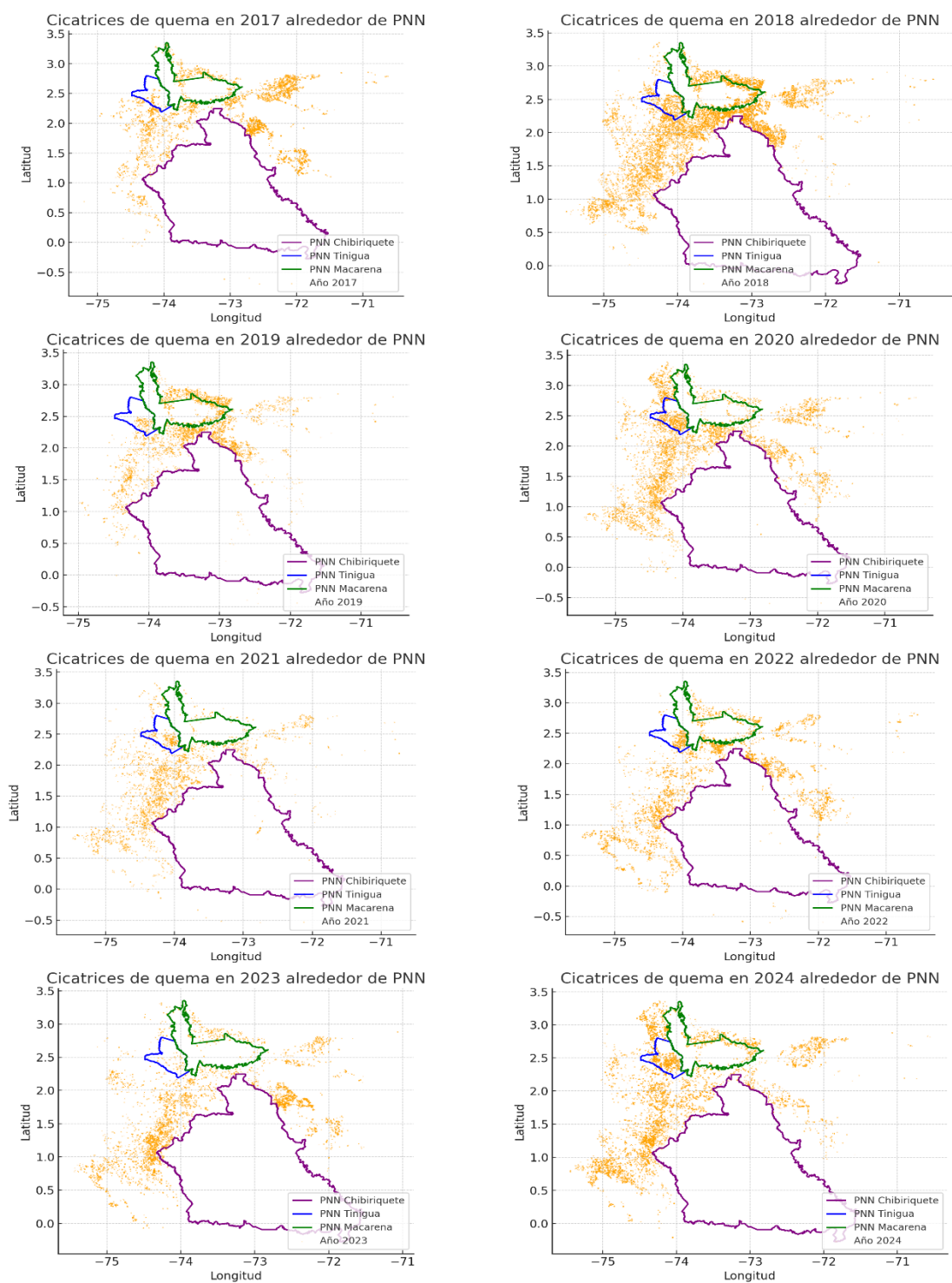
Las cicatrices de quema, que representan áreas afectadas por incendios, muestran una distribución espacial similar a los puntos de calor, confirmando su relación como indicadores de la degradación forestal. Los años 2018, 2020 y 2022 presentan una mayor cantidad de cicatrices, sugiriendo incendios importantes en esos periodos, destacándose especialmente la proximidad de las cicatrices al Parque Tinigua. El año 2023 presenta una leve reducción para un repunte, de nuevo, en eventos en el año 2024.

Los boxplots de las distancias de las cicatrices de quema por municipio revelan patrones importantes. En municipios como San Vicente del Caguán y San José del Guaviare, se observa una gran cantidad de valores atípicos, lo cual indica la presencia de eventos extremos en cuanto a la distancia de las cicatrices respecto al parque. Estos valores sugieren que, aunque muchas cicatrices de quema ocurren a distancias moderadas, existen eventos donde los incendios afectan zonas muy cercanas o significativamente alejadas del parque, aumentando la complejidad en el manejo del riesgo de deforestación. En el caso del municipio de Solano, se destaca una menor dispersión en las distancias, lo que podría implicar una concentración más consistente de los eventos de quema en ciertas áreas.

Además, la variabilidad en las distancias de las cicatrices al parque sugiere que los incendios no siempre están localizados cerca de los límites del parque, sino que la afectación se distribuye en un área más amplia.

Figura 6

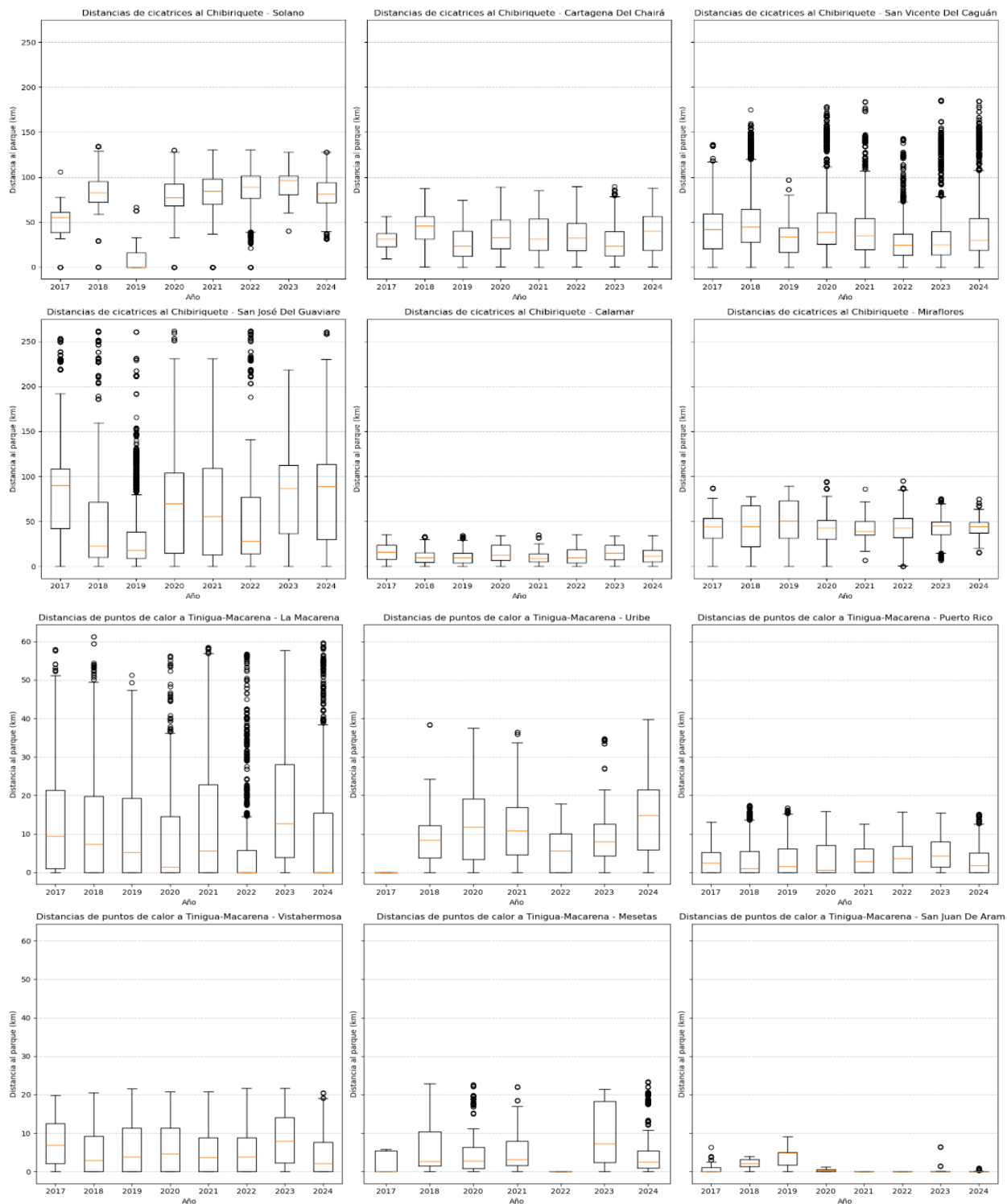
Comportamiento Temporal Cicatrices de Quema en PNN



Nota. Comportamiento cicatrices de quema alrededor de PNN.

Figura 7

Boxplots Distancias de Cicatrices a PNN



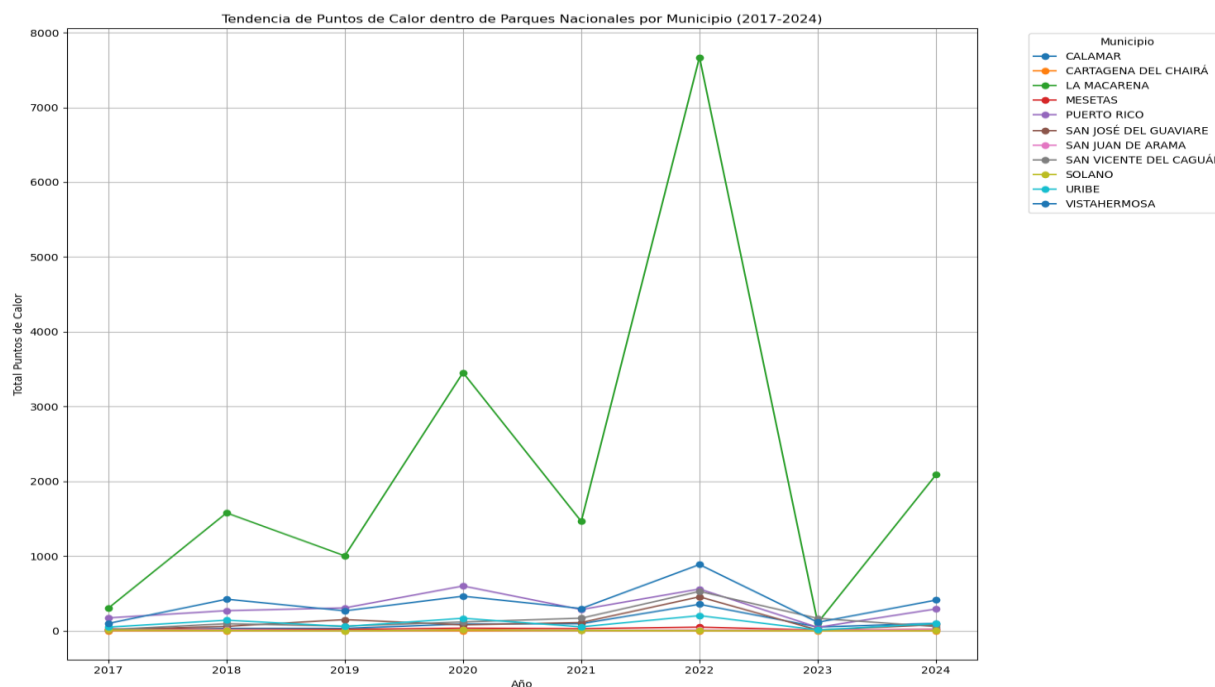
Nota. Visualización acercamiento de cicatrices de quema sobre PNN.

Tendencias de Puntos de Calor dentro de PNN

La tendencia de puntos de calor dentro de PNN por municipio muestra un pico notable en el año 2018, lo cual podría estar relacionado con factores climáticos o actividades humanas que favorecieron la ocurrencia de incendios o eventos de deforestación. Municipios como La Macarena y Puerto Rico y Vistahermosa destacan por su alta actividad de puntos de calor en este periodo. Para el año 2024 se identifica un incremento en las tendencias para todos los municipios, especialmente para el municipio de la Macarena.

Figura 8

Tendencia Puntos de Calor 2017-2024



Nota. Visualización tendencia de puntos de calor dentro de PNN.

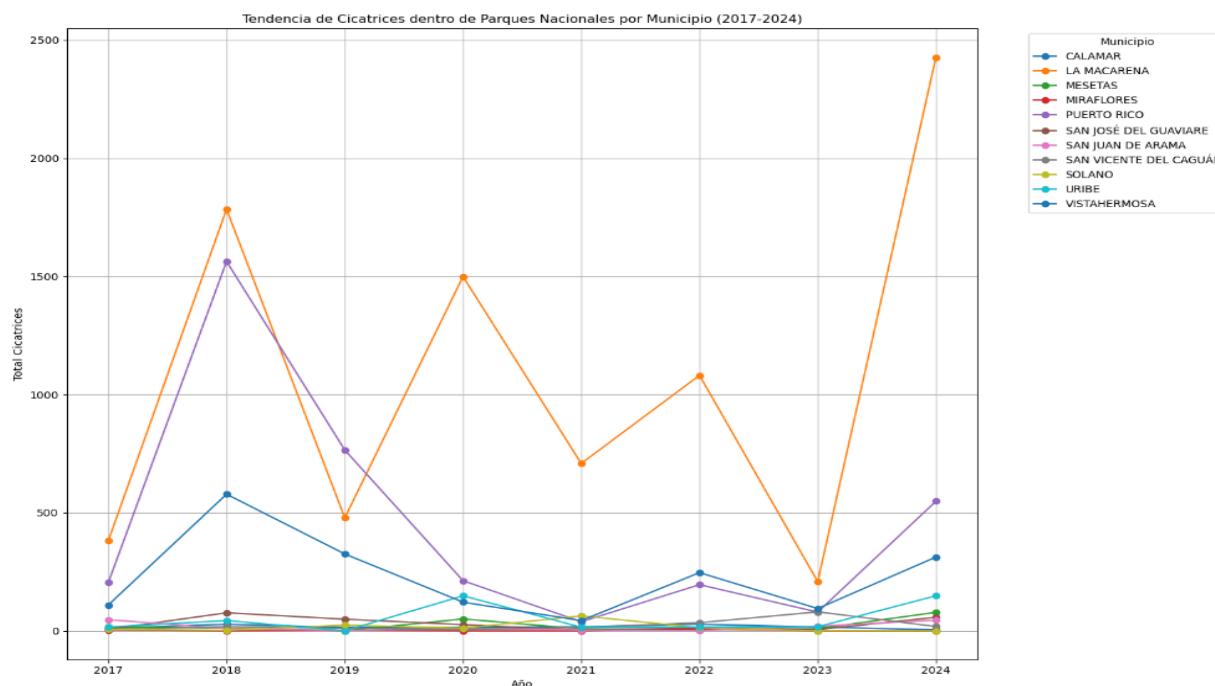
Tendencias de Cicatrice de Quema dentro de PNN

Las tendencias de cicatrices de quema presentan picos más pronunciados para el año 2022, lo que indica que este fue un año crítico para la degradación forestal. El municipio de La

Macarena nuevamente es el que presente el comportamiento de mayor actividad, lo cual lo convierte en una zona especialmente vulnerable.

Figura 9

Tendencia Cicatrices de Quema 2017-2024



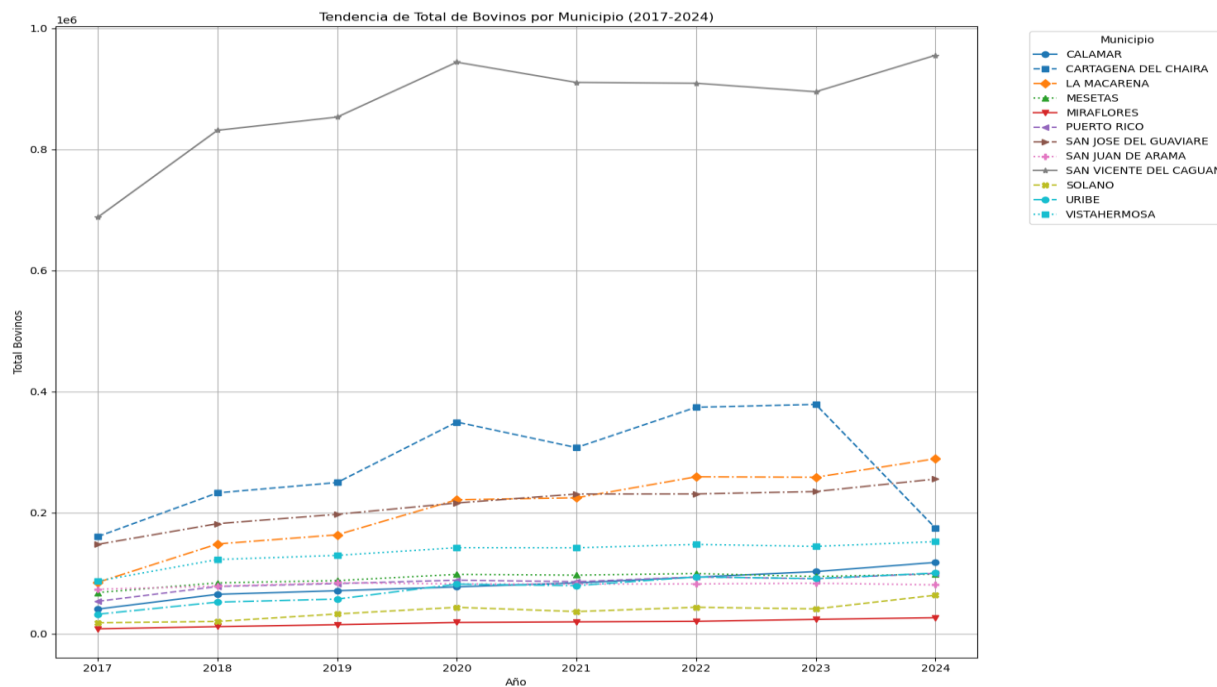
Nota. Visualización tendencia de cicatrices de quema dentro de PNN.

Tendencia Cabezas de Ganado

El análisis de la tendencia de cabezas de ganado revela un crecimiento sostenido en la mayoría de los municipios a lo largo del periodo 2017-2024. Municipios como San Vicente del Caguán, Cartagena del Chairá y La Macarena han registrado los incrementos más notables, lo cual es preocupante debido a la relación entre la ganadería extensiva y la deforestación. Este aumento sostenido sugiere una presión constante y creciente sobre el ecosistema de los parques.

Figura 10

Tendencia Cabezas de Ganado 2017-2024



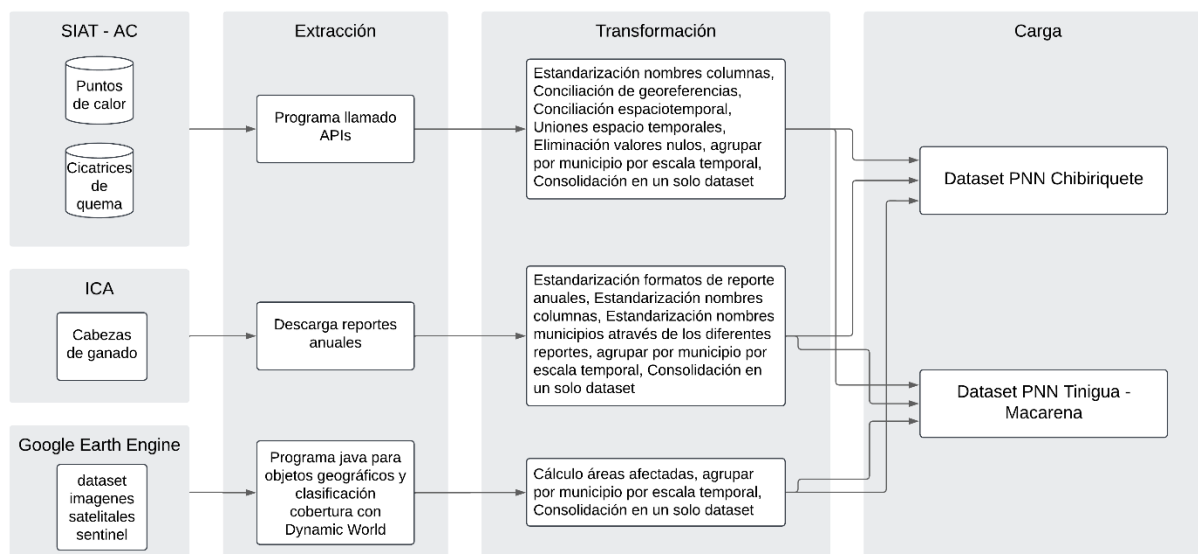
Nota. Visualización tendencia Cabezas de Ganado por Municipio.

Extracción, Transformación y Carga de Datos

En esta fase se realizan las actividades de extracción de la información correspondiente a la conexión a base de datos por medio de APIs para los puntos de calor y cicatrices de quema, descargar datos de fuentes como el ICA, el IGAC y el DANE. También, a la generación de datos de clasificación de cobertura de suelo y descarga de datos georreferenciados desde la plataforma de GEE.

Figura 11

Flujo ETL



Nota. Pasos para la extracción, transformación y carga de los datos requeridos para el desarrollo del estudio.

Las transformaciones están relacionadas con:

Estándarización de Datos: Unificación de sistemas de coordenadas y formatos, asegurando la correcta integración de datos geospaciales. Esto incluye la transformación de datos geográficos a sistemas de referencia homogéneos que permitan su integración precisa.

Limpieza de Datos: Identificación y eliminación de valores nulos, eliminación de columnas irrelevantes y normalización de los nombres de municipios para garantizar la consistencia. Este proceso es fundamental para asegurar que los datos utilizados sean de la más alta calidad y que no existan sesgos derivados de datos faltantes o erróneos.

Interpolación Temporal: Ajuste de las series temporales para alinear las frecuencias de las distintas variables, facilitando un análisis coherente a nivel cuatrimestral. Esta alineación temporal es esencial para realizar comparaciones consistentes y para identificar tendencias a lo largo del tiempo.

Finalmente, se cargan los diferentes datos en archivos tipo geojson y csv. Se crea finalmente dos conjuntos de datos mediante los cuales se realiza los análisis de variables.

Puntos de Calor

Los datos de puntos de calor se obtuvieron de los sensores satelitales MODIS y VIIRS, a través de una API proporcionada por el Instituto Amazónico de Investigaciones Científicas (SINCHI) para los puntos de calor. Los datos son georreferenciados y para cada punto de calor se relaciona su ubicación específica mediante una geometría tipo punto. La frecuencia de los datos es diaria. Se recolectaron más de 270,000 registros, abarcando el periodo de enero de 2017 a agosto de 2024, se realizan transformaciones, filtraciones y agrupaciones requeridas y se cargan los datos en los siguientes archivos para los análisis posteriores:

- “puntosdecalor_2017_2024.geojson”
- “puntoscalor_mensualizados.csv”
- “puntoscalor_mensualizados_pivot.csv”

Cicatrices de Quema

Los datos de cicatrices de quema fueron recolectados mediante imágenes satelitales de Sentinel-2 y Landsat 8, procesados con el Índice de Nube Diferenciada (NBR) y a través de la API proporcionada por el Instituto Amazónico de Investigaciones Científicas (SINCHI). La frecuencia de los datos es mensual, obteniéndose alrededor de 258,000 registros distribuidos en los municipios de estudio para el periodo temporal del 2017 a agosto de 2024. Estos datos son georreferenciados y permiten identificar las áreas afectadas por el fuego, proporcionando información crucial para comprender los patrones de quema. Los datos de cicatrices de quema no tienen referencia a los municipios en las que se detectan. Por lo tanto, con el objeto de preparar el dataset para los análisis requeridos se obtienen las georreferencias de los municipios de interés y se practica una unión espacial con el conjunto de datos de cicatrices de quema empleando geopandas. Se realizan transformaciones, filtraciones y agrupaciones requeridas y se cargan los datos en los siguientes archivos:

- “cquemmas_2017_2024.geojson”
- “cicatrices_mensualizados.csv”
- “cicatrices_mensualizados_pivot.csv”

Explotación Ganadera

Se utilizaron registros oficiales de censos de bovinos de 2017 a 2024 proporcionados por el ICA. Estos datos están en formato xlsx con periodicidad anual. Se filtran los datos por los departamentos de Meta, Caquetá y Guaviare. Los datos fueron preprocesados, eliminando columnas irrelevantes y normalizando los nombres de los municipios para mantener la consistencia. Se concatenan los dataframes para obtener un dataset consolidado. La información

de la explotación ganadera es esencial, ya que esta actividad es uno de los principales motores de la deforestación en la región.

Los datos son cargados en el archivo: “consolidado_ganado.csv”

Praderización en PNN

Para obtener los datos de praderización en los parques nacionales se toma como referencia los municipios de influencia en el área protegida de interés. Luego, se realiza en la plataforma de GEE una intersección de la geometría del parque nacional y de la geometría de cada uno de los municipios. Sobre la nueva geometría se aplica la clasificación de cobertura de suelo (lc) de Dynamic World. Esta clasificación se determina con una frecuencia cuatrimestral desde 2017 a agosto de 2024. Igualmente, se calcula en GEE la cantidad de píxeles para las categorías de interés: “grass” y “shrub and scrub” que posteriormente se convierten en área empleando la escala proporcionada por las imágenes sentinel 2 que son de 10m. Estas áreas representan las zonas que son praderas o están en proceso de praderización.

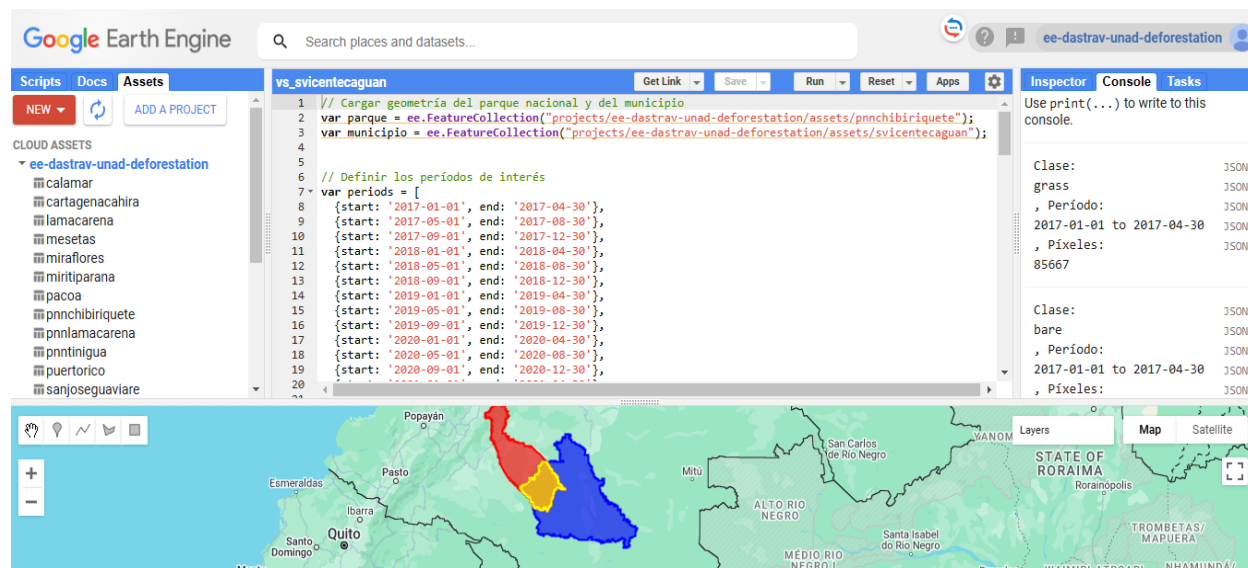
Posteriormente, los datos son procesados en pandas para agregar columnas de identificación de los municipios, dar formato a las fechas, agrupar los datos por municipio y calcular las áreas de praderas.

Los datos son cargados en los archivos:

- “chibiriquete_consolidado.csv”
- “tinigua_macarena_consolidado.csv”

Figura 12

Plataforma de Google Earth Engine.



Nota. Proceso de cálculo de áreas de praderización para el pnn chibiriquete con relación al municipio San Vicente del Caguán.

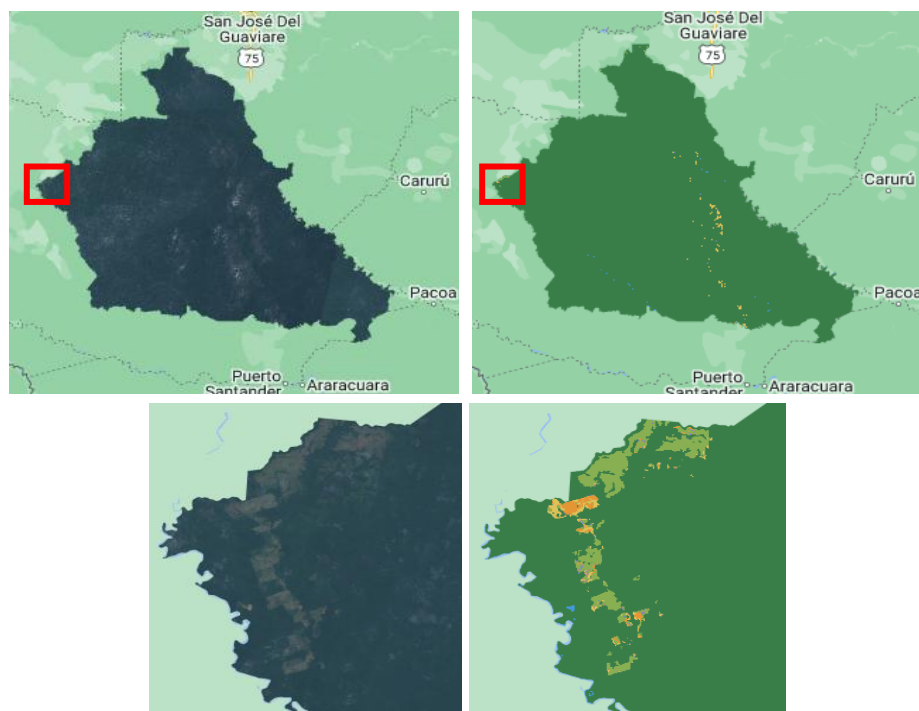
Dataset Consolidado

En esta etapa se realiza, mediante operaciones de agrupamiento (“merge”) en pandas, la integración de los archivos csv creados en las etapas de extracción y transformación anterior.

Los datos se cargan en los archivos:


1. “dataset_chibiriquete_2017_2024.csv”
2. “dataset_tinigua_macarena_2017_2024.csv”

Se presenta en la Figura 14 una muestra de los datasets finales. En este caso para el parque nacional tinigua_macarena.

Figura 13*Clasificación de Coberturas con Dynamic World*

Nota. Se muestra imagen Sentinel 2 del chibiriquete con clasificación de Pradera (verde claro).

Figura 14*Dataset PNN Tinigua_Macarena*

	Municipio	Año_Cuatrimestre	Puntos de Calor	Cicatrices de Quema	Cabezas de Ganado	Degradación_bosque (ha)
0	CALAMAR	2017-Q1	662	2285.0	40733.0	489.0
1	CALAMAR	2017-Q2	76	216.0	40733.0	479.0
2	CALAMAR	2017-Q3	102	6.0	40733.0	560.0
3	CALAMAR	2018-Q1	1448	2483.0	65077.0	755.0
4	CALAMAR	2018-Q2	340	172.0	65077.0	168.0
...
133	SAN VICENTE DEL CAGUAN	2023- 	2295	1793.0	895290.0	346.0
134	SAN VICENTE DEL CAGUAN	2023-Q2	308	406.0	895290.0	624.0
135	SAN VICENTE DEL CAGUAN	2023-Q3	237	2217.0	895290.0	1511.0
136	SAN VICENTE DEL CAGUAN	2024-Q1	3448	5041.0	955534.0	974.0
137	SAN VICENTE DEL CAGUAN	2024-Q2	280	993.0	955534.0	927.0

138 rows × 6 columns

Nota. La figura muestra el resultado final de la etapa ETL para el parque tinigua_macarena.

Análisis Exploratorio de Datos

El análisis exploratorio de datos (EDA) es una etapa fundamental para comprender las relaciones y patrones inherentes a las variables involucradas en el fenómeno de deforestación en los Parques Nacionales Natural Serranía del Chibiriquete y Tingua-Macarena. A continuación, se presenta un análisis detallado basado en los histogramas y boxplots de cuatro variables clave: degradación del bosque, puntos de calor, cicatrices de quema, y cabezas de ganado. Este análisis proporciona una perspectiva inicial sobre la distribución y la variabilidad de los datos, permitiendo identificar tendencias, valores atípicos y potenciales correlaciones.

En la Figura 15 se presentan las visualizaciones de los histogramas para el PNN Chibiriquete.

Distribución de la Degradación del Bosque PNN Chibiriquete

El histograma muestra que la mayor parte de los valores de degradación del bosque se concentran en áreas pequeñas (menores a 500 ha), mientras que una minoría significativa alcanza valores muy elevados, con algunos municipios exhibiendo más de 5000 ha de degradación. Esta distribución, con una fuerte asimetría positiva, sugiere la presencia de eventos extremos de degradación en ciertas localidades, lo cual podría estar vinculado a la conversión del uso del suelo para actividades como ganadería extensiva o expansión de cultivos. La presencia de valores extremos en la distribución refuerza la necesidad de abordar estos eventos de forma diferenciada, dado que representan focos críticos de intervención.

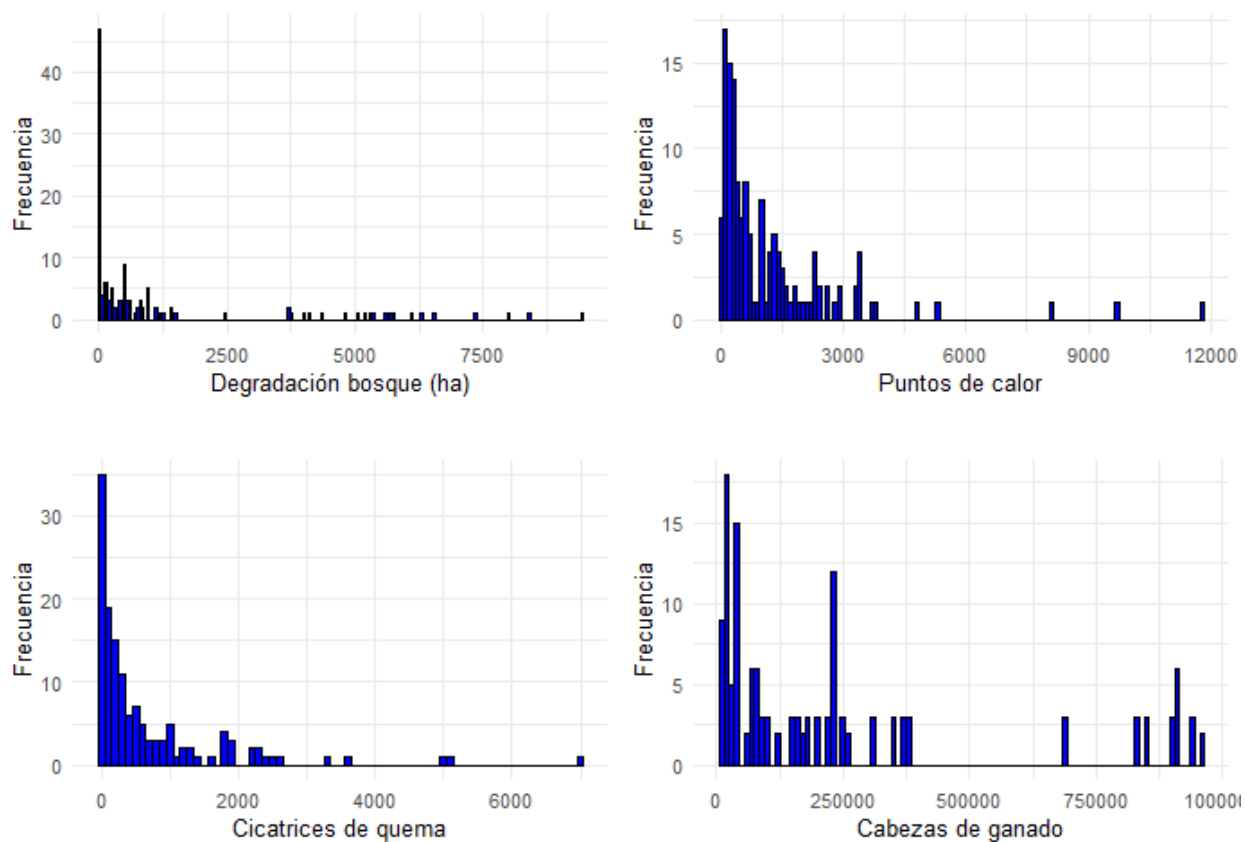
Distribución de los Puntos de Calor PNN Chibiriquete

El análisis de la distribución de los puntos de calor revela una tendencia similar a la de la degradación del bosque: una alta concentración de valores bajos y una cola larga hacia la derecha. La mayoría de las observaciones presentan pocos puntos de calor, mientras que algunas

superan los 3000. Esta concentración indica que la ocurrencia de eventos de alta intensidad de calor es menos frecuente, pero significativa, y puede estar asociada a incendios forestales o actividades de quema agrícola descontrolada. Estos puntos de calor son un importante indicador de áreas con riesgo elevado de degradación forestal, sugiriendo que los esfuerzos de conservación deberían enfocarse en estos picos extremos.

Figura 15

Distribución de las Variables PNN Chibiriquete



Nota. La figura muestra el comportamiento de distribución de los factores de riesgo en PNN Chibiriquete.

Distribución de las Cicatrices de Quema PNN Chibiriquete

Al igual que los puntos de calor, las cicatrices de quema muestran una distribución asimétrica, con la mayoría de los valores inferiores a 1000 ha, pero con algunos casos extremos que alcanzan más de 2000 ha. Las cicatrices de quema reflejan el área total afectada por incendios y, al igual que los puntos de calor, presentan una variabilidad considerable.

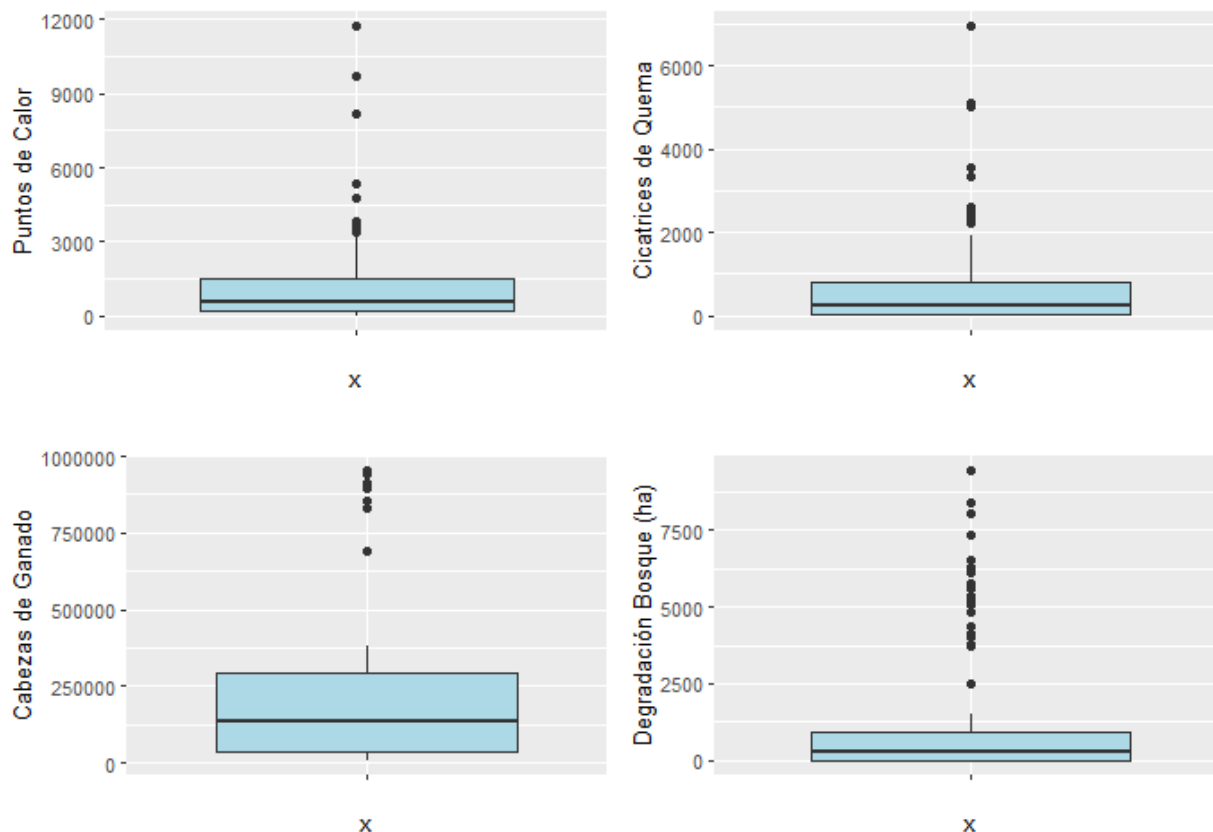
Distribución de las Cabezas de Ganado PNN Chibiriquete

El análisis de la distribución de cabezas de ganado evidencia un comportamiento bimodal. Existe un grupo significativo de municipios con menos de 250,000 cabezas de ganado, mientras que otro grupo alcanza valores superiores a 750,000 cabezas. Esta variabilidad sugiere la coexistencia de dos tipos de productores: pequeños y grandes ganaderos. La alta concentración de ganado en ciertos municipios puede ser un fuerte motor de deforestación, ya que la ganadería extensiva requiere grandes extensiones de tierra. La identificación de estas áreas permite diseñar estrategias específicas de manejo y mitigación, enfocadas en los municipios con mayor presión ganadera.

Análisis de Boxplots PNN Chibiriquete

El análisis de los boxplots complementa la información presentada en los histogramas, proporcionando una visualización clara de la distribución de los datos y de los valores atípicos.

En la Figura 16 se presentan las gráficas de boxplots respectivas.

Figura 16*Boxplots Variables PNN Chibiriquete*

Nota. Exploración de las distribuciones y presencia de los Outliers en PNN Chibiriquete.

Boxplot de Puntos de Calor

La mayoría de los puntos de calor se encuentran en un rango relativamente bajo, con una mediana cercana a cero. Sin embargo, existen valores atípicos que alcanzan hasta 12,000 puntos, indicando eventos anómalos de alta actividad que podrían estar asociados a incendios extensivos. Estos eventos sugieren la necesidad de implementar sistemas de alerta temprana que permitan una intervención rápida en áreas afectadas.

Boxplot de Cicatrices de Quema

Similar a los puntos de calor, la mayoría de las observaciones de cicatrices de quema son relativamente bajas, pero los valores atípicos alcanzan hasta 6000 ha. Estos eventos extremos reflejan incendios forestales significativos, cuyas causas podrían estar relacionadas tanto con actividades antropogénicas como con condiciones climáticas adversas. El manejo de estos valores extremos es crucial para mitigar el impacto sobre la biodiversidad y evitar la expansión de áreas quemadas.

Boxplot de Cabezas de Ganado

La mediana de cabezas de ganado se sitúa alrededor de 250,000, con un rango intercuartil amplio. La presencia de varios valores atípicos superiores a 750,000 cabezas subraya la concentración desigual de la ganadería en la región. Esto plantea desafíos específicos en términos de uso del suelo y requiere intervenciones enfocadas a promover prácticas ganaderas sostenibles que minimicen la presión sobre los ecosistemas forestales.

Boxplot de la Degradación del Bosque

La mayoría de las observaciones de degradación del bosque son bajas, con una mediana próxima a cero. Sin embargo, los valores atípicos alcanzan más de 7500 ha, lo cual es indicativo de eventos severos de degradación que no se distribuyen de manera uniforme. Estos eventos extremos son los principales impulsores de la pérdida de cobertura en la región y merecen especial atención en las estrategias de conservación.

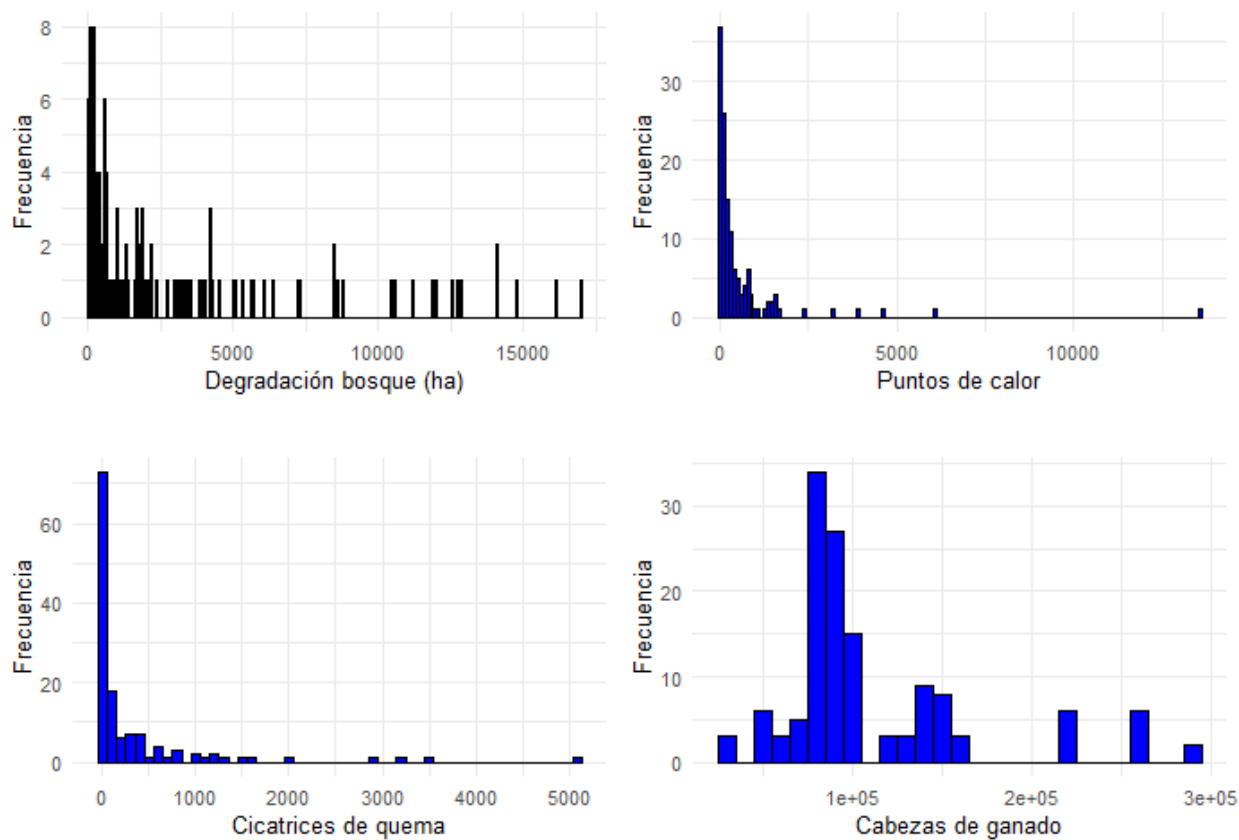
Los gráficos analizados evidencian una alta variabilidad en todas las variables estudiadas, destacando la presencia de valores atípicos significativos en varias de ellas. Estos valores extremos indican que existen eventos particulares o áreas específicas que están sufriendo

impactos desproporcionadamente altos, como la degradación del bosque, puntos de calor intensos y cicatrices de quema.

Distribución de la Degradación del Bosque PNN Tinigua-Macarena

Figura 17

Distribución de las Variables PNN Tinigua-Macarena



Nota. La figura muestra el comportamiento de distribución de los factores de riesgo en PNN Tinigua-Macarena.

En el histograma, se observa una distribución fuertemente sesgada hacia la derecha, con la mayoría de los valores concentrados en los rangos más bajos de degradación del bosque, y unos pocos valores extremos hacia la derecha. Esto sugiere que, en la mayoría de los casos, la

degradación del bosque es baja, pero existen algunas observaciones con niveles de degradación significativamente altos, lo cual puede ser indicativo de eventos puntuales o de actividades más intensas de deforestación.

Distribución de las Cicatrices de Quema PNN Tinigua-Macarena

El histograma muestra una distribución con un sesgo a la derecha, donde la mayoría de las observaciones se encuentran cercanas al valor cero, lo que indica que hay pocos eventos de quema en la mayoría de las áreas. Sin embargo, existen algunas observaciones con valores elevados, lo que sugiere que ciertos sectores han sido mucho más afectados por el fuego.

Distribución de las Cabezas de Ganado PNN Tinigua-Macarena

La distribución de las cabezas de ganado muestra una alta concentración en los valores más bajos, aunque la dispersión es algo mayor que en los otros gráficos. El histograma tiene varias "mesetas" con picos en diferentes intervalos, lo cual sugiere la existencia de diferencias marcadas en la cantidad de ganado entre diferentes zonas o comunidades.

Análisis de Boxplots PNN Tinigua-Macarena

En la Figura 18 se presentan los boxplots de las variables para el Parque Nacional Natural Tinigua-Macarena.

Boxplot de Puntos de Calor

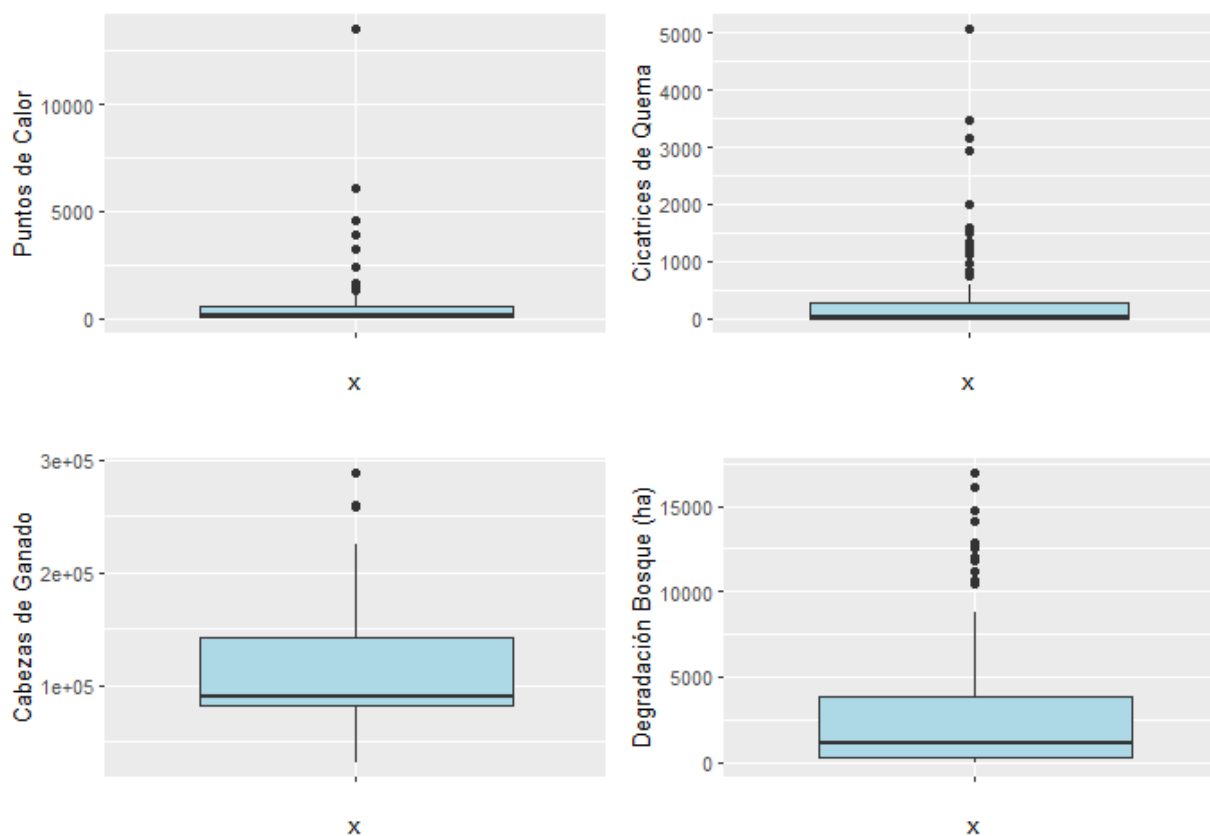
La distribución de los puntos de calor muestra un sesgo significativo hacia valores bajos, con una cantidad considerable de puntos atípicos. Esto indica que la mayoría de las áreas tienen baja presencia de puntos de calor, pero existen ciertas zonas con niveles extremadamente altos, lo cual podría estar asociado a eventos de incendios o actividades humanas intensivas.

Boxplot de Cicatrices de Quema

Similar al análisis de los puntos de calor, el boxplot muestra que la mayoría de las observaciones tienen valores bajos, pero hay muchos outliers que reflejan áreas con alta intensidad de quema. Este patrón destaca la importancia de analizar estos valores atípicos, ya que pueden representar eventos anómalos o acciones intencionadas.

Figura 18

Boxplots Variables PNN Tinigua-Macarena



Nota. Exploración de las distribuciones y presencia de los Outliers en PNN Tinigua-Macarena.

Boxplot de Cabezas de Ganado

En el caso de las cabezas de ganado, se observa una mayor variabilidad en los datos, con una mediana alta y presencia de algunos valores atípicos. Este comportamiento sugiere que la cría de ganado es una actividad relativamente común y extendida en la región, con algunas áreas que presentan una mayor concentración de ganado.

Boxplot de la Degradación del Bosque

La degradación del bosque presenta un comportamiento similar al de las cicatrices de quema y los puntos de calor, con una concentración de la mayoría de los valores en el rango bajo y la presencia de valores atípicos. La distribución sugiere que la degradación significativa está concentrada en unas pocas áreas, lo cual es importante para identificar los focos de intervención prioritaria.

Relaciones entre Variables para el PNN Chibiriquete

Para profundizar en la comprensión de los factores que influyen en la degradación del bosque en el Parque Nacional Natural Serranía del Chibiriquete, se ha llevado a cabo un análisis de regresión lineal múltiple. Este análisis tiene como objetivo determinar la relación entre la degradación del bosque (variable dependiente) y varios factores de riesgo identificados, tales como los puntos de calor, cicatrices de quema, cabezas de ganado, año cuatrimestre y la ubicación geográfica representada por los municipios (variables independientes). A continuación, se presenta la metodología adoptada, así como los resultados del análisis y su interpretación en el contexto del fenómeno de deforestación.

Transformación Logarítmica y Escalado

Antes de proceder al análisis de regresión lineal múltiple, fue necesario realizar ajustes en los datos debido a las características específicas de sus distribuciones. Como se observó en el análisis exploratorio de datos (EDA), las variables involucradas presentan distribuciones altamente sesgadas, con colas largas hacia valores elevados. Este comportamiento se refleja en los histogramas que muestran una concentración considerable de valores bajos y una menor frecuencia de valores extremos.

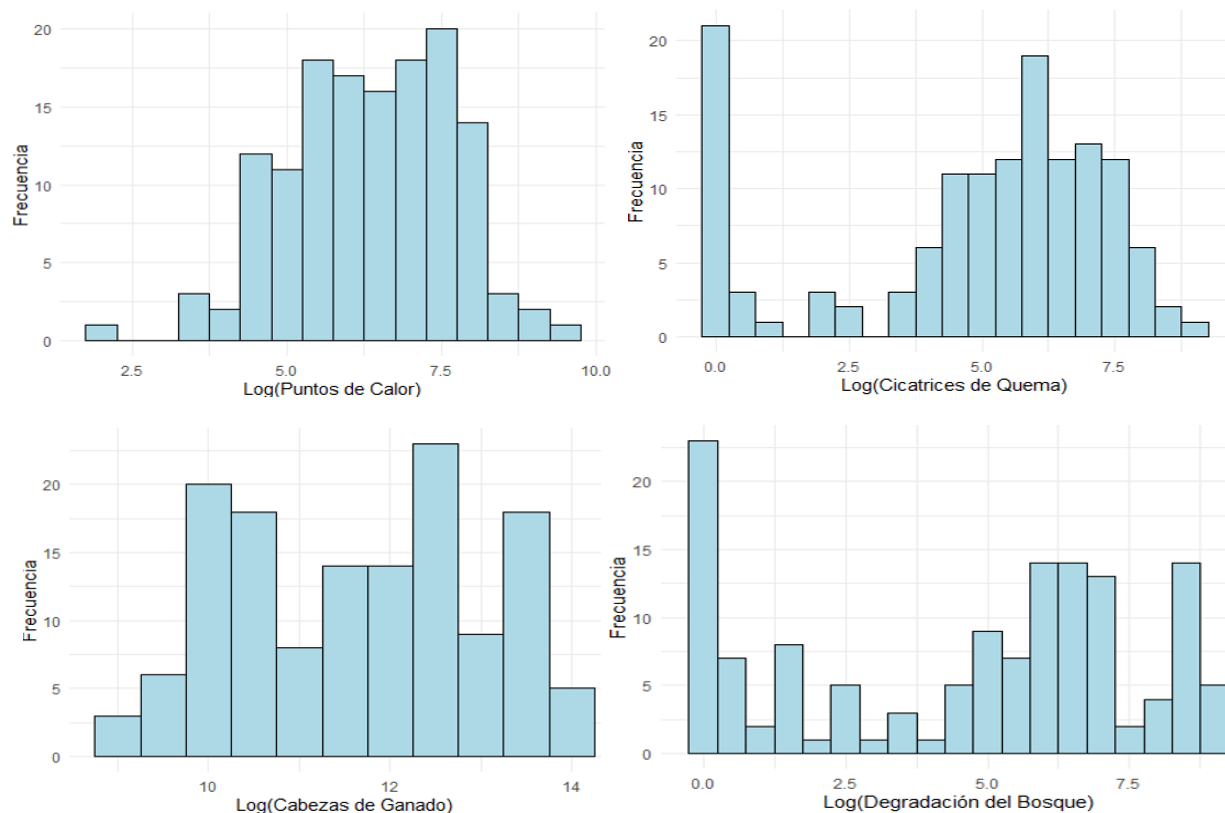
Debido a esta asimetría y para manejar las escalas disímiles de las variables, se aplicaron dos transformaciones clave:

Transformación Logarítmica: Todas las variables independientes (puntos de calor, cicatrices de quema, cabezas de ganado) fueron transformadas mediante una función logarítmica. Esta transformación tiene como objetivo reducir la asimetría y estabilizar la varianza, haciendo que las distribuciones sean más adecuadas para el análisis de regresión (Grigorev Alexey, 2021).

Sin embargo, tras la transformación logarítmica, las distribuciones no lograron alcanzar formas cercanas a una normal.

Figura 19

Distribuciones Suavizadas



Nota. Visualización del efecto de la transformación logarítmica sobre las distribuciones.

Escalado Min-Max: Posteriormente, se utilizó el método de escalado min-max para llevar todas las variables transformadas a un rango de $[0, 1]$. Este procedimiento se adoptó para asegurar que las variables fueran comparables y que ninguna de ellas dominara la regresión debido a la disparidad en sus escalas originales. El escalado min-max es particularmente útil en modelos que son sensibles a las escalas de los datos, como lo es la regresión lineal, al permitir que los coeficientes sean interpretados de manera más uniforme.

Modelo de Regresión Lineal Múltiple

La regresión lineal múltiple fue utilizada para evaluar la relación entre la degradación del bosque (variable dependiente) y los factores explicativos previamente descritos. La ecuación general del modelo puede representarse de la siguiente manera:

- $$\text{Degradación del Bosque} = \beta_0 + \beta_1 \text{Puntos de Calor} + \beta_2 \text{Cicatrices de Quema} + \beta_3 \text{Cabezas de Ganado} + \beta_4 \text{Año_Cuatrimestre} + \beta_5 \text{Municipio} + \epsilon$$

Los resultados se describen a continuación y pueden observarse en la Figura 20.

Coefficiente de determinación (R^2): El valor de R^2 es 0.9778, lo cual indica que aproximadamente el 97.78% de la variabilidad en la degradación del bosque puede ser explicada por las variables incluidas en el modelo. Este alto valor sugiere que el modelo tiene un buen ajuste a los datos. El valor ajustado R^2_{ajustado} es de 0.9716, lo que confirma que el modelo sigue siendo robusto incluso al considerar el número de variables.

Prueba F: La estadística F es de 157.4, con un valor p asociado de $< 2.2e-16$, lo cual indica que el modelo es significativo en general, lo que implica que al menos una de las variables independientes tiene un efecto significativo sobre la degradación del bosque.

Coefficientes del Modelo

A continuación, se destacan los coeficientes más relevantes, su significancia estadística y la interpretación de los resultados:

Intercepto: El intercepto tiene un valor de 0.805, lo que representa la estimación promedio de la degradación del bosque en ausencia de cualquier efecto de las variables independientes.

Figura 20

Resultado Regresión Lineal Múltiple

```
summary(mlr_model_suavizado)
## Call:
## lm(formula = scaled_Degradacion_Bosque ~ Municipio + Año_Cuatrimstre +
##     scaled_Puntos_Calor + scaled_Cicatrices_Quema + scaled_Cabezas_Ganado,
##     data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.203629 -0.028045  0.002764  0.032761  0.113756
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.805230   0.071440  11.271 < 2e-16 ***
## MunicipioCARTAGENA DEL CHAIRA -0.571421   0.043934 -13.006 < 2e-16 ***
## MunicipioMIRAFLORES -0.634534   0.052146 -12.168 < 2e-16 ***
## MunicipioSAN JOSE DEL GUAVIARE -0.085860   0.036261  -2.368  0.01969 *
## MunicipioSAN VICENTE DEL CAGUAN  0.236205   0.079922  2.955  0.00384 **
## MunicipioSOLANO  0.203078   0.031241  6.500 2.61e-09 ***
## Año_Cuatrimstre2017-Q2 -0.066320   0.037194 -1.783  0.07741 .
## Año_Cuatrimstre2017-Q3 -0.033546   0.036225 -0.926  0.35651
## Año_Cuatrimstre2018-Q1  0.029979   0.033795  0.887  0.37704
## Año_Cuatrimstre2018-Q2 -0.091600   0.034993 -2.618  0.01014 *
## Año_Cuatrimstre2018-Q3 -0.017048   0.034806 -0.490  0.62528
## Año_Cuatrimstre2019-Q1  0.007701   0.035833  0.215  0.83024
## Año_Cuatrimstre2019-Q2 -0.020220   0.043037 -0.470  0.63943
## Año_Cuatrimstre2019-Q3  0.027281   0.043538  0.627  0.53225
## Año_Cuatrimstre2020-Q1  0.008093   0.038732  0.209  0.83489
## Año_Cuatrimstre2020-Q2 -0.004620   0.043863 -0.105  0.91630
## Año_Cuatrimstre2020-Q3  0.052636   0.039384  1.336  0.18422
## Año_Cuatrimstre2021-Q1  0.051110   0.039086  1.308  0.19380
## Año_Cuatrimstre2021-Q2 -0.006324   0.044115 -0.143  0.88629
## Año_Cuatrimstre2021-Q3  0.053902   0.040234  1.340  0.18317
## Año_Cuatrimstre2022-Q1  0.039796   0.042357  0.940  0.34957
## Año_Cuatrimstre2022-Q2  0.011819   0.042269  0.280  0.78032
## Año_Cuatrimstre2022-Q3  0.079049   0.039864  1.983  0.04993 *
## Año_Cuatrimstre2023-Q1  0.037290   0.040080  0.930  0.35427
## Año_Cuatrimstre2023-Q2  0.038923   0.047004  0.828  0.40947
## Año_Cuatrimstre2023-Q3  0.085537   0.046563  1.837  0.06898 .
## Año_Cuatrimstre2024-Q1  0.035153   0.040349  0.871  0.38559
## Año_Cuatrimstre2024-Q2  0.018051   0.044530  0.405  0.68602
## scaled_Puntos_Calor  0.032912   0.073395  0.448  0.65475
## scaled_Cicatrices_Quema  0.007914   0.027968  0.283  0.77775
## scaled_Cabezas_Ganado -0.360193   0.156159 -2.307  0.02300 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05603 on 107 degrees of freedom
## Multiple R-squared:  0.9778, Adjusted R-squared:  0.9716
## F-statistic: 157.4 on 30 and 107 DF,  p-value: < 2.2e-16
```

Nota. Este es el resumen de la ejecución del modelo de regresión lineal múltiple que permite evaluar las relaciones entre las variables.

Municipios: Varios municipios presentaron coeficientes negativos significativos, como Cartagena del Chairá (-0.571 , $p < 2 \times 10^{-16}$) y Miraflores (-0.634 , $p < 2 \times 10^{-16}$). Esto indica que, en comparación con el municipio de referencia (Calamar), estas ubicaciones experimentan una

degradación del bosque significativamente menor. Sin embargo, San Vicente del Caguán y Solano mostraron coeficientes positivos significativos, sugiriendo una mayor degradación.

Año - cuatrimestre: La variable temporal muestra variabilidad en los coeficientes, con algunos cuatrimestres presentando efectos significativos. Por ejemplo, el segundo cuatrimestre de 2018 tiene un coeficiente negativo (-0.0916 , $p = 0.010$), lo cual indica una disminución en la degradación durante ese periodo. Esto podría estar relacionado con variaciones estacionales o intervenciones específicas durante esos años.

Puntos de calor : La variable puntos de calor no resultó ser significativa ($p=0.654$), lo cual sugiere que, una vez escalada y transformada, su contribución a la degradación del bosque no es tan importante como se esperaba inicialmente. Esto podría indicar una alta colinealidad con otras variables del modelo.

Cicatrices de quema: Similar a los puntos de calor, las cicatrices de quema no mostraron una significancia estadística relevante ($p=0.778$), lo que podría sugerir que su relación con la degradación del bosque es más compleja de lo que puede captar un modelo lineal.

Cabezas de ganado: La variable cabezas de ganado presentó un coeficiente negativo significativo (-0.360 , $p = 0.023$). Este resultado es contraintuitivo, ya que la literatura reporta a la ganadería como un impulsor clave de la deforestación (Indepaz, 2024). Este comportamiento podría reflejar una relación no lineal entre la ganadería y la degradación del bosque.

Análisis de residuos: Los residuos del modelo presentan un rango relativamente estrecho (mínimo: -0.204 , máximo: 0.114), lo que sugiere que el modelo ajusta los datos en su mayoría, aunque no sin excepciones. La presencia de algunos patrones residuales no aleatorios sugiere que el modelo lineal podría no capturar por completo la complejidad del fenómeno, indicando la necesidad de explorar modelos no lineales o técnicas más avanzadas.

Los resultados del análisis de regresión lineal múltiple proporcionan una perspectiva valiosa sobre los factores que influyen en la degradación del bosque en el Parque Nacional Serranía del Chiriquete. Sin embargo, se identificaron varias limitaciones:

Significancia de inconsistente de variables clave: A pesar de que los puntos de calor y las cicatrices de quema son conocidos como factores importantes de deforestación, no resultaron significativos en el modelo lineal. Esto podría deberse a la colinealidad o a una relación más compleja con la degradación que no es lineal.

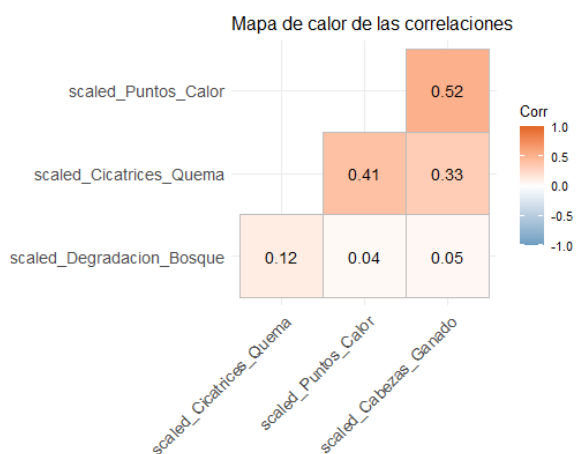
Relación contraintuitiva con la ganadería: El coeficiente negativo asociado a las cabezas de ganado sugiere que la dinámica de la ganadería en relación con la deforestación es más complicada de lo que se puede modelar con una regresión lineal simple.

Análisis de Correlación

Para evaluar la relación entre las variables cuantitativas, se realizó un análisis de correlación y se generó un mapa de calor para visualizar mejor las relaciones.

Figura 21

Correlación Variables para PNN Chiriquete



Nota. Visualización del nivel de correlación existente para factores de riesgo para el PNN Tinigua-Macarena.

La baja correlación entre la degradación del bosque y las variables independientes indica que las relaciones entre estas variables podrían ser más complejas de lo que el análisis de correlación lineal puede captar. Es posible que existan interacciones no lineales entre los factores, o bien efectos acumulativos y sinérgicos que requieren un modelo más flexible para ser identificados adecuadamente.

Modelo Aditivo Generalizado

Para superar las limitaciones observadas en el modelo de regresión lineal múltiple, particularmente en cuanto a la capacidad de capturar relaciones no lineales entre los factores de riesgo y la degradación del bosque, se decidió aplicar un Modelo Aditivo Generalizado (GAM). Los modelos GAM son útiles cuando se sospecha que las relaciones entre las variables predictoras y la respuesta no son estrictamente lineales, como es el caso de los procesos ambientales complejos. Los GAM permiten modelar relaciones no lineales utilizando funciones suavizadas (splines), lo cual ofrece una mayor flexibilidad para capturar patrones complejos en los datos (James et al., 2017).

En la Figura 22 se presentan los resultados del modelo GAM.

Coefficientes Suavizados

Puntos de calor: La estimación del término suavizado no resultó significativa ($p=0.244$), lo cual indica que la relación no lineal entre los puntos de calor y la degradación del bosque no es lo suficientemente fuerte como para ser considerada significativa en este modelo. Esto podría sugerir que los puntos de calor, tal y como están representados, no explican de forma contundente la degradación observada.

Cicatrices de quema: El término suavizado correspondiente a las cicatrices de quema tiene un valor de p significativo ($p=0.0059$), con un edf (grados de libertad efectivos) de 1.066,

indicando que hay una relación no lineal que explica una parte importante de la degradación del bosque. Esto sugiere que, aunque la relación no sea compleja, es relevante.

Figura 22

Resultado Modelo GAM

```
# Cargar Librerías necesarias
library(mgcv)

# Convertir Municipio a factor
df$Municipio <- as.factor(df$Municipio)

# Creación del modelo GAM
modelo_gam <- gam(scaled_Degradacion_Bosque ~ s(scaled_Puntos_Calor) + s(scaled_Cicatrices_Quema) + s(scaled_Cabezas_Ganado) + Municipio,
                  data = df)

# Resumen del modelo GAM
summary(modelo_gam)

## Family: gaussian
## Link function: identity
##
## Formula:
## scaled_Degradacion_Bosque ~ s(scaled_Puntos_Calor) + s(scaled_Cicatrices_Quema) +
##   s(scaled_Cabezas_Ganado) + Municipio
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.75142    0.03713  20.237 < 2e-16 ***
## MunicipioCARTAGENA DEL CHAIRA -0.82705    0.05576 -14.831 < 2e-16 ***
## MunicipioMIRAFLORES          -0.54344    0.03695 -14.707 < 2e-16 ***
## MunicipioSAN JOSE DEL GUAVIARE -0.33002    0.05631  -5.861 3.97e-08 ***
## MunicipioSAN VICENTE DEL CAGUAN -0.04474    0.13073  -0.342  0.733
## MunicipioSOLANO              0.25214    0.02807   8.981 3.91e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##              edf Ref.df    F p-value
## s(scaled_Puntos_Calor)  1.000  1.000  1.371 0.24397
## s(scaled_Cicatrices_Quema) 1.066  1.128  7.038 0.00594 **
## s(scaled_Cabezas_Ganado)  7.287  8.122  3.447 0.00159 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) = 0.972  Deviance explained = 97.5%
## GCV = 0.0034764  Scale est. = 0.0030896  n = 138
```

Nota. Resultados del modelo aditivo generalizado que permite evaluar la relevancia de las variables.

Cabezas de ganado: Esta variable presenta una relación compleja con la degradación del bosque, con un edf de 7.287 y un valor de p muy significativo ($p=0.0016$). Esto indica una

relación no lineal considerable, lo cual puede estar reflejando los diferentes modos en que la actividad ganadera impacta el bosque.

R-cuadrado ajustado (R^2): El modelo muestra un R^2 ajustado de 0.972, lo cual indica que el 97.2% de la variabilidad en la degradación del bosque está explicada por el modelo. Este es un ajuste significativamente alto, lo cual sugiere que el modelo GAM está capturando de manera adecuada los patrones de degradación.

Validación cruzada generalizada (GCV): El valor de GCV es 0.0035, indicando que el error de predicción del modelo es bajo, lo cual es un buen indicador de la capacidad del modelo para generalizar en nuevos datos.

Evaluación de los Resultados (Validación Cruzada)

Una vez se ha establecido un modelo para la interpretación y el estudio del fenómeno de la deforestación con base en las variables independientes, se emplea la técnica estadística de la validación cruzada con el objeto de evaluar la idoneidad del modelo. Esta técnica proporciona una evaluación confiable de la capacidad del modelo para hacer predicciones en nuevos datos. De esta manera es posible estimar el rendimiento del modelo de manera más robusta, asegurando que los resultados no dependan únicamente de una partición específica de los datos.

En la Figura 23 se presentan los resultados de la validación cruzada del modelo GAM. Dichos resultados revelan diferencias entre el MSE, MAE y R^2 obtenidos en el conjunto de prueba y los valores correspondientes en la validación cruzada. Estas discrepancias sugieren la presencia de sobreajuste, lo que indica que, aunque el modelo muestra un buen desempeño durante la validación cruzada, tiene dificultades para generalizar a nuevos conjuntos de datos. Esto implica que el modelo podría estar capturando detalles específicos de los datos de entrenamiento en lugar de aprender patrones más generales del fenómeno de deforestación.

Figura 23

Validación Cruzada del Modelo GAM

```

library(mgcv)

# Dividir Los datos en conjunto de entrenamiento (80%) y prueba (20%)
set.seed(123)
trainIndex <- sample(seq_len(nrow(df)), size = 0.8 * nrow(df))
trainData <- df[trainIndex, ]
testData <- df[-trainIndex, ]

# Entrenar el modelo GAM en Los datos de entrenamiento
modelo_gam <- gam(scaled_Degradacion_Bosque ~ s(scaled_Puntos_Calor) + s(scaled_Cicatrices_Quema) + s(scaled_Cabezas_Ganado) + Municipio,
                  data = trainData)

# Predicción en Los datos de prueba
predicciones_gam_test <- predict(modelo_gam, testData)

# Evaluar el rendimiento en el conjunto de prueba
mse_gam_test <- mean((testData$scaled_Degradacion_Bosque - predicciones_gam_test)^2)
rsq_gam_test <- cor(testData$scaled_Degradacion_Bosque, predicciones_gam_test)^2
mae_gam_test <- mean(abs(testData$scaled_Degradacion_Bosque - predicciones_gam_test))

cat("MSE en prueba: ", mse_gam_test, "\n")
## MSE en prueba: 0.006654462
cat("R-Squared en prueba: ", rsq_gam_test, "\n")
## R-Squared en prueba: 0.9376828
cat("MAE en prueba: ", mae_gam_test, "\n")
## MAE en prueba: 0.05863225

# Validación cruzada manual k-fold
# Remover espacios o caracteres especiales en la variable Municipio
df$Municipio <- gsub(" ", "_", df$Municipio)
df$Municipio <- as.factor(df$Municipio)

# Función para realizar validación cruzada manual k-fold
k_fold_cv <- function(data, k = 10) {
  n <- nrow(data)
  folds <- sample(rep(1:k, length.out = n))

  mse_list <- numeric(k)
  mae_list <- numeric(k)
  rsq_list <- numeric(k)

  for(i in 1:k) {
    # Dividir en entrenamiento y prueba
    test_indices <- which(folds == i)
    test_data <- data[test_indices, ]
    train_data <- data[-test_indices, ]

    # Entrenar el modelo GAM
    model_gam <- gam(scaled_Degradacion_Bosque ~ s(scaled_Puntos_Calor) + s(scaled_Cicatrices_Quema) + s(scaled_Cabezas_Ganado) + Municipio,
                    data = train_data)

    # Predicciones
    pred_test <- predict(model_gam, test_data)

    # Calcular métricas para este fold
    mse_list[i] <- mean((test_data$scaled_Degradacion_Bosque - pred_test)^2)
    mae_list[i] <- mean(abs(test_data$scaled_Degradacion_Bosque - pred_test))
    rsq_list[i] <- cor(test_data$scaled_Degradacion_Bosque, pred_test)^2
  }
}

```

```

# Devolver Las métricas promedio
list(
  mse = mean(mse_list),
  mae = mean(mae_list),
  rsq = mean(rsq_list)
)
}

# Ejecutar La validación cruzada
cv_results <- k_fold_cv(df, k = 10)

cat("MSE con validación cruzada manual: ", cv_results$mse, "\n")
## MSE con validación cruzada manual: 0.003771342
cat("MAE con validación cruzada manual: ", cv_results$mae, "\n")
## MAE con validación cruzada manual: 0.04703198
cat("R-Squared con validación cruzada manual: ", cv_results$rsq, "\n")
## R-Squared con validación cruzada manual: 0.9684705

```

Nota. Resultados de la validación cruzada para el modelo GAM.

Buscando reducir el sobreajuste, se procede a ajustar los parámetros de suavizado del modelo GAM con el objetivo de optimizar su rendimiento. En los modelos GAM, el ajuste clave radica en la "suavidad" de las funciones spline que se emplean para modelar los términos no lineales. En el enfoque implementado mediante la librería mgcv, esta suavidad se controla a través de los parámetros "k" (número de bases de spline) y " λ " (penalización de suavidad).

Se implementa el ajuste de estos parámetros mediante una búsqueda en rejilla (grid search), optimizando el valor de "k" para regular la flexibilidad de las splines y utilizando validación cruzada para evaluar el rendimiento del modelo ajustado. En la Figura 24 se muestran los resultados del modelo con el ajuste del parámetro "k".

A pesar de los esfuerzos por mejorar el modelo mediante estos ajustes, los resultados indican que no se consigue una mejora significativa en su rendimiento. Para optimizar aún más el modelo, sería necesario contar con una mayor cantidad de datos o eliminar variables que no aporten información relevante. Sin embargo, estas alternativas no son viables en el presente caso.

Figura 24

Validación Cruzada con Ajuste de Parámetros del Modelo GAM

```

library(mgcv)

# Dividir los datos en conjunto de entrenamiento (80%) y prueba (20%)
set.seed(123)
trainIndex <- sample(seq_len(nrow(df)), size = 0.8 * nrow(df))
trainData <- df[trainIndex, ]
testData <- df[-trainIndex, ]

# Crear función para ajuste de hiperparámetros mediante búsqueda en rejilla
tune_gam <- function(data, k_values, folds = 5) {
  # Lista para almacenar los resultados
  tuning_results <- list()

  # Generar los folds para validación cruzada
  set.seed(123)
  n <- nrow(data)
  fold_indices <- sample(rep(1:folds, length.out = n))

  # Iterar sobre los valores de k para ajustar el modelo
  for (k in k_values) {
    mse_list <- numeric(folds)

    for (i in 1:folds) {
      # Dividir en entrenamiento y prueba
      test_indices <- which(fold_indices == i)
      test_data <- data[test_indices, ]
      train_data <- data[-test_indices, ]

      # Entrenar el modelo GAM con el valor de k específico
      model_gam <- gam(scaled_Degradacion_Bosque ~ s(scaled_Cicatrices_Quema, k = k) +
                      s(scaled_Cabezas_Ganado, k = k) + Municipio,
                      data = train_data)

      # Predicciones
      pred_test <- predict(model_gam, test_data)

      # Calcular MSE para este fold
      mse_list[i] <- mean((test_data$scaled_Degradacion_Bosque - pred_test)^2)
    }

    # Guardar el MSE promedio para este valor de k
    tuning_results[[paste0("k=", k)] <- mean(mse_list)
  }

  # Devolver el resultado con el menor MSE
  best_k <- names(which.min(unlist(tuning_results)))
  cat("Mejor valor de k:", best_k, "\n")
  return(tuning_results)
}

# Definir los valores de k a probar (ajuste de suavidad)
k_values <- c(3, 5, 7, 9, 12)

# Ejecutar la búsqueda en rejilla para ajustar k
tuning_results <- tune_gam(trainData, k_values, folds = 5)

## Mejor valor de k: k=12

# Entrenar el modelo final con el mejor valor de k
best_k <- as.numeric(gsub("k=", "", names(which.min(unlist(tuning_results)))))
modelo_gam_tuned <- gam(scaled_Degradacion_Bosque ~ s(scaled_Puntos_Calor, k = best_k) +
                      s(scaled_Cicatrices_Quema, k = best_k) +

```

```

      s(scaled_Cabezas_Ganado, k = best_k) + Municipio,
      data = trainData)

# Evaluar el rendimiento del modelo ajustado en el conjunto de prueba
predicciones_gam_test <- predict(modelo_gam_tuned, testData)
mse_gam_test <- mean((testData$scaled_Degradacion_Bosque - predicciones_gam_test)^2)
rsq_gam_test <- cor(testData$scaled_Degradacion_Bosque, predicciones_gam_test)^2
mae_gam_test <- mean(abs(testData$scaled_Degradacion_Bosque - predicciones_gam_test))

cat("MSE en prueba (modelo ajustado): ", mse_gam_test, "\n")
## MSE en prueba (modelo ajustado): 0.006421528
cat("R-Squared en prueba (modelo ajustado): ", rsq_gam_test, "\n")
## R-Squared en prueba (modelo ajustado): 0.9401211
cat("MAE en prueba (modelo ajustado): ", mae_gam_test, "\n")
## MAE en prueba (modelo ajustado): 0.05605367

```

Nota. Resultados del modelo después de ajuste parámetros. *Fuente.* Autor

No obstante, aunque los resultados en el conjunto de prueba son inferiores a los de la validación cruzada, la magnitud de la diferencia no es crítica. Adicionalmente, el modelo ajustado sigue explicando el 94% de la variabilidad en la degradación del bosque, lo cual lo convierte en una herramienta suficientemente válida tanto para la modelación del fenómeno como para la construcción de un Indicador de Riesgo de Deforestación (IRD) aplicable al Parque Nacional Natural Serranía del Chiriquete.

Relaciones entre Variables para el PNN Tinigua-Macarena

De manera similar que para el Parque Natural Nacional Chiriquete se realizará una transformación logarítmica y escalado min-max. de las variables y se procede a la implementación de una regresión lineal múltiple.

Modelo de Regresión Lineal Múltiple

Se presenta a continuación los resultados de la regresión para el PNN Tinigua-Macarena.

Figura 25

Modelo de Regresión Lineal Múltiple para PNN Tinigua-Macarena

```
summary(mlr_model_suavizado)##
## Call:
## lm(formula = scaled_Degradacion_Bosque ~ Municipio + Año_Cuatrimestre +
##     scaled_Puntos_Calor + scaled_Cicatrices_Quema + scaled_Cabezas_Ganado,
##     data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.34632 -0.02945 -0.00523  0.04296  0.17492
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.046470   0.157589   6.641 1.55e-09 ***
## MunicipioMESETAS -0.438203   0.074872  -5.853 5.91e-08 ***
## MunicipioPUERTO RICO -0.152799   0.061534  -2.483 0.01465 *
## MunicipioSAN JUAN DE ARAMA -0.576292   0.114140  -5.049 1.95e-06 ***
## MunicipioURIBE -0.368283   0.073434  -5.015 2.24e-06 ***
## MunicipioVISTAHERMOSA -0.193673   0.048119  -4.025 0.00011 ***
## Año_Cuatrimestre2017-Q2 -0.058653   0.058357  -1.005 0.31723
## Año_Cuatrimestre2017-Q3 -0.002514   0.061776  -0.041 0.96762
## Año_Cuatrimestre2018-Q1  0.031401   0.052764   0.595 0.55307
## Año_Cuatrimestre2018-Q2 -0.030284   0.059089  -0.513 0.60940
## Año_Cuatrimestre2018-Q3  0.029125   0.054794   0.532 0.59620
## Año_Cuatrimestre2019-Q1 -0.116103   0.054660  -2.124 0.03608 *
## Año_Cuatrimestre2019-Q2 -0.002558   0.071071  -0.036 0.97136
## Año_Cuatrimestre2019-Q3  0.017940   0.075124   0.239 0.81174
## Año_Cuatrimestre2020-Q1  0.012601   0.059977   0.210 0.83401
## Año_Cuatrimestre2020-Q2 -0.036085   0.067360  -0.536 0.59333
## Año_Cuatrimestre2020-Q3  0.072975   0.061739   1.182 0.23996
## Año_Cuatrimestre2021-Q1  0.058883   0.059387   0.992 0.32378
## Año_Cuatrimestre2021-Q2 -0.163336   0.063243  -2.583 0.01122 *
## Año_Cuatrimestre2021-Q3 -0.100467   0.065040  -1.545 0.12552
## Año_Cuatrimestre2022-Q1 -0.020011   0.065132  -0.307 0.75929
## Año_Cuatrimestre2022-Q2  0.052355   0.072217   0.725 0.47013
## Año_Cuatrimestre2022-Q3  0.095455   0.062134   1.536 0.12757
## Año_Cuatrimestre2023-Q1  0.075690   0.063753   1.187 0.23789
## Año_Cuatrimestre2023-Q2  0.013287   0.077978   0.170 0.86504
## Año_Cuatrimestre2023-Q3  0.006956   0.079634   0.087 0.93056
## Año_Cuatrimestre2024-Q1  0.022020   0.062802   0.351 0.72659
## Año_Cuatrimestre2024-Q2  0.042407   0.064478   0.658 0.51221
## scaled_Puntos_Calor -0.104793   0.163702  -0.640 0.52351
## scaled_Cicatrices_Quema  0.001587   0.052305   0.030 0.97586
## scaled_Cabezas_Ganado -0.035594   0.128871  -0.276 0.78295
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.08287 on 102 degrees of freedom
## Multiple R-squared: 0.8592, Adjusted R-squared: 0.8178
## F-statistic: 20.76 on 30 and 102 DF, p-value: < 2.2e-16
```

Nota. Resumen del modelo de regresión lineal múltiple.

Coefficiente de determinación (R^2): El valor del R^2 múltiple es de 0.8592, lo cual sugiere que el modelo es capaz de explicar el 85.92% de la variabilidad de la degradación del bosque en el Parque Nacional Tinigua-Macarena. Este es un valor relativamente alto, lo cual indica que las variables incluidas en el modelo tienen un buen poder explicativo sobre la degradación del bosque. Sin embargo, dado que el R^2 ajustado es 0.8178, se observa una leve disminución, lo que indica que algunos de los predictores pueden no estar aportando significativamente al modelo, particularmente teniendo en cuenta la cantidad de variables incluidas.

Prueba F: La prueba F tiene un valor estadístico de 20.76 con un p-valor menor a $2.2e-16$, lo cual indica que el modelo es globalmente significativo. Esto significa que, en conjunto, las variables independientes explican de manera significativa la variación en la degradación del bosque. Es decir, al menos una de las variables incluidas en el modelo tiene un efecto significativo sobre la variable dependiente.

Coeficientes del Modelo

Intercepto: El valor del intercepto es 1.046470, con un p-valor altamente significativo ($1.55e-09$). Esto sugiere que, cuando todas las variables predictoras se encuentran en su nivel base, la degradación del bosque tiene un valor base positivo.

Municipios: Las variables correspondientes a los municipios muestran efectos significativos en varios casos.

Los municipios Mesetas, San Juan de Arama, Uribe y Vista Hermosa tienen coeficientes negativos significativos, lo cual indica que en estos municipios la degradación del bosque es menor en comparación con el municipio de referencia (Calamar).

Año cuatrimestre: La mayoría de las categorías del factor Año Cuatrimestre no muestran una significancia estadística notable. Esto implica que el tiempo del año (por cuatrimestre) no tiene un efecto claro sobre la degradación del bosque, al menos con los datos disponibles y bajo este modelo. Algunas excepciones son 2019-Q1 y 2021-Q2, que presentan efectos significativos, sugiriendo que estos períodos específicos tuvieron un impacto en la degradación.

Puntos de calor : El coeficiente de `scaled_Puntos_Calor` es negativo (-0.104793) y no significativo ($p = 0.52351$). Esto resulta sorprendente dado que los puntos de calor generalmente se asocian con una mayor incidencia de incendios y deforestación. La falta de significancia puede deberse a varios factores, como una correlación alta con otras variables incluidas en el modelo (multicolinealidad) o a una alta variabilidad de la variable de puntos de calor que no se explica adecuadamente con el modelo lineal.

Cicatrices de quema: El coeficiente de `scaled_Cicatrices_Quema` es prácticamente cero (0.001587) y también no significativo ($p = 0.97586$). Este resultado sugiere que, en el contexto del modelo, la variable cicatrices de quema no contribuye significativamente a explicar la variabilidad en la degradación del bosque. Esto podría ser el resultado de la colinealidad con otras variables o de la baja variabilidad en la variable de cicatrices en comparación con la variable dependiente.

Cabezas de ganado: El coeficiente de `scaled_Cabezas_Ganado` es también negativo (-0.035594) y no significativo ($p = 0.78295$). Este resultado es contrario a lo esperado, ya que la presencia de ganado generalmente se asocia con un mayor impacto sobre la deforestación debido

a la conversión de áreas forestales en pastizales. La falta de significancia podría estar indicando que la relación entre ganado y degradación no es lineal o que existen otros factores no considerados en el modelo que afectan esta relación.

Análisis de residuos: Los residuos presentan un mínimo de -0.34632 y un máximo de 0.17492, con una mediana cercana a cero (-0.00523), lo cual sugiere que los residuos están razonablemente bien centrados alrededor de cero, aunque parece haber cierta asimetría en los extremos negativos.

La desviación estándar de los residuos (Residual standard error: 0.08287) indica una variabilidad moderada, y la alta concentración de residuos cerca del cero sugiere un buen ajuste global del modelo.

Significancia de inconsistente de variables clave: La no significancia de las variables puntos de calor, cicatrices de quema y cabezas de ganado podría indicar que estas variables no tienen un impacto directo linealmente sobre la degradación del bosque en el Parque Nacional Tinigua-Macarena en este modelo específico. Esto podría deberse a:

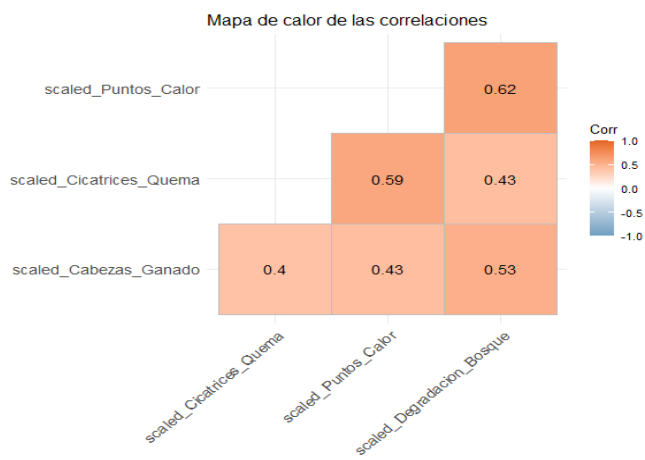
- **Multicolinealidad:** Si estas variables están altamente correlacionadas entre sí, el modelo no es capaz de distinguir sus efectos individuales de manera efectiva.
- **Relaciones no lineales:** Las relaciones podrían ser no lineales, y un modelo de regresión lineal múltiple no es adecuado para capturar estas interacciones.

Análisis de Correlación

Para evaluar la relación entre las variables cuantitativas, se realizó un análisis de correlación y se generó un mapa de calor para visualizar mejor las relaciones. Los valores de correlación obtenidos son los siguientes:

Figura 26

Correlación Variables para PNN Tinigua-Macarena



Nota. Visualización del nivel de correlación existente para factores de riesgo para el PNN Tinigua-Macarena.

scaled_Puntos_Calor y scaled_Degradacion_Bosque presentan una correlación de 0.62. Esto indica una relación moderada-alta entre los puntos de calor y la degradación del bosque, lo cual sugiere que a medida que aumentan los puntos de calor, la degradación del bosque también tiende a aumentar.

scaled_Cicatrices_Quema y scaled_Puntos_Calor tienen una correlación de 0.59, lo cual indica que estas dos variables están relacionadas y probablemente representan eventos conectados. Es decir, donde se observan más puntos de calor, es probable que se produzcan más cicatrices de quema.

scaled_Cabezas_Ganado y scaled_Degradacion_Bosque tienen una correlación de 0.53, también indicando una relación positiva moderada entre la explotación ganadera y la degradación del bosque.

Las correlaciones entre algunas de las variables independientes son altas (como 0.59 entre `scaled_Puntos_Calor` y `scaled_Cicatrices_Quema`). Esto puede llevar a problemas de multicolinealidad, lo cual afectaría la capacidad del modelo para estimar correctamente los coeficientes de estas variables.

Regularización Ridge

Se aplica a continuación la regularización Ridge que puede ayudar a reducir la multicolinealidad de los factores de riesgo penalizando los coeficientes altos y llevarlos a valores más estables.

Este método puede hacer que los coeficientes sean menos sensibles a las correlaciones con otras variables.

En la Figura 27 se observa el modelo planteado y los resultados obtenidos.

Coefficientes del Modelo Ridge

Intercepto: El valor del intercepto es 0.681. Este valor representa el valor predicho de la variable dependiente cuando todas las variables independientes son iguales a cero. Sugiere que existe un nivel base de degradación del bosque incluso sin la influencia de otras variables.

Municipios: Los municipios presentan diferentes coeficientes, indicando variabilidad en la degradación del bosque entre distintas localidades. Los municipios San Juan de Arama, Mesetas y Uribe tienen coeficientes negativos notables, lo que sugiere una menor degradación del bosque en estos lugares en comparación con el intercepto.

Puntos de calor: El coeficiente para `scaled_Puntos_Calor` es 0.072, lo cual indica que, con otras variables controladas, un aumento en los puntos de calor se asocia con un ligero aumento en la degradación del bosque. Aunque la magnitud de este coeficiente no es muy alta, sigue siendo positivo, sugiriendo una relación directa.

Cicatrices de quema: El coeficiente es bastante bajo (0.003), lo que indica que la influencia de cicatrices de quema en la degradación del bosque es mínima en comparación con otras variables. Esto podría deberse a una alta correlación con otras variables que afectan la deforestación.

Figura 27

Modelo Ridge para PNN Tinigua-Macarena

```
# Variables independientes a matriz
x <- model.matrix(scaled_Degradacion_Bosque ~ Municipio + scaled_Puntos_Calor + scaled_Cicatrices_Quema +
scaled_Cabezas_Ganado, data = df)[, -1]

# Variable dependiente
y <- df$scaled_Degradacion_Bosque

# Modelo Ridge (alpha = 0 para Ridge)
ridge_model <- cv.glmnet(x, y, alpha = 0, standardize = TRUE)

# Coeficientes del modelo Ridge usando Lambda óptima
ridge_coefs <- coef(ridge_model, s = "lambda.min")
print(ridge_coefs)

## 9 x 1 sparse Matrix of class "dgCMatrix"
##              s1
## (Intercept)    0.681072165
## MunicipioMESETAS -0.243218715
## MunicipioPUERTO RICO    0.009758209
## MunicipioSAN JUAN DE ARAMA -0.325799670
## MunicipioURIBE    -0.168829820
## MunicipioVISTAHERMOSA -0.071344417
## scaled_Puntos_Calor    0.072309203
## scaled_Cicatrices_Quema    0.003132642
## scaled_Cabezas_Ganado    0.208779974

# Predicciones para el modelo Ridge usando Lambda óptima
ridge_predictions <- predict(ridge_model, s = "lambda.min", newx = x)

# Calcular el error estándar residual para Ridge
ridge_rse <- sqrt(mean((y - ridge_predictions)^2))
print(paste("Error estándar residual (RSE) para Ridge:", ridge_rse))

## [1] "Error estándar residual (RSE) para Ridge: 0.0999365714059608"

# Calcular R^2 para Ridge
ss_total <- sum((y - mean(y))^2) # Suma total de Los cuadrados
ss_residual <- sum((y - ridge_predictions)^2) # Suma de Los residuos al cuadrado
ridge_r2 <- 1 - (ss_residual / ss_total) # Coeficiente de determinación R^2
print(paste("R^2 para Ridge:", ridge_r2))

## [1] "R^2 para Ridge: 0.733073152434138"
```

Nota. Resumen de resultados del modelo Ridge para PNN Tinigua-Macarena.

Error Estándar Residual (RSE)

El error estándar residual (RSE) obtenido es 0.0999, lo cual indica que el modelo tiene un error promedio bajo al predecir los valores de la variable dependiente escalada. Este valor implica una buena capacidad del modelo para ajustarse a los datos observados, considerando que la escala es de 0 a 1.

Coefficiente de Determinación (R^2)

El coeficiente de determinación R^2 obtenido es 0.733, lo cual sugiere que el 73.3% de la variabilidad en la degradación del bosque se explica por el modelo Ridge. Esto indica un ajuste aceptable del modelo, ya que la mayor parte de la variabilidad de la degradación del bosque está siendo explicada por las variables consideradas.

Evaluación de los Resultados (Validación Cruzada)

De manera similar a como se realizó con el modelo GAM se practica la validación cruzada para el modelo Ridge planteado para el estudio de la deforestación en el parque nacional Tinigua-Macarena.

Los resultados se muestran en la siguiente figura.

Figura 28

Validación Cruzada para el Modelo Ridge

```
library(glmnet)

# Preparar Los datos para La regresión Ridge
x <- model.matrix(scaled_Degradacion_Bosque ~ Municipio + scaled_Puntos_Calor + scaled_Cicatrices_Quema +
scaled_Cabezas_Ganado, data = df)[, -1]
y <- df$scaled_Degradacion_Bosque

# Definir parámetros para La validación cruzada
set.seed(123) # Asegura La reproducibilidad
k <- 10 # Número de folds

# Crear folds
folds <- sample(rep(1:k, length.out = length(y)))

# Inicializar listas para almacenar errores y métricas
rmse_list <- numeric(k)
mae_list <- numeric(k)
rsq_list <- numeric(k)
```

```

# Realizar la validación cruzada k-fold
for (i in 1:k) {
  # Dividir los datos en conjunto de entrenamiento y prueba
  train_indices <- which(folds != i)
  test_indices <- which(folds == i)

  x_train <- x[train_indices, ]
  y_train <- y[train_indices]

  x_test <- x[test_indices, ]
  y_test <- y[test_indices]

  # Ajustar el modelo Ridge en el conjunto de entrenamiento (alpha = 0 para Ridge)
  ridge_model <- glmnet(x_train, y_train, alpha = 0, standardize = TRUE)

  # Obtener Lambda óptima usando validación cruzada interna
  best_lambda <- cv.glmnet(x_train, y_train, alpha = 0)$lambda.min

  # Realizar predicciones en el conjunto de prueba
  predictions <- predict(ridge_model, s = best_lambda, newx = x_test)

  # Calcular métricas: RMSE, MAE, y R2
  residuals <- y_test - predictions
  rmse_list[i] <- sqrt(mean(residuals^2)) # RMSE
  mae_list[i] <- mean(abs(residuals)) # MAE
  rsq_list[i] <- 1 - (sum(residuals^2) / sum((y_test - mean(y_test))^2)) # R2
}

# Calcular métricas promedio de validación cruzada
ridge_cv_rmse <- mean(rmse_list)
ridge_cv_mae <- mean(mae_list)
ridge_cv_rsqa <- mean(rsq_list)

# Mostrar resultados
cat("Cross-Validation RMSE for Ridge Model:", ridge_cv_rmse, "\n")
## Cross-Validation RMSE for Ridge Model: 0.1022899
cat("Cross-Validation MAE for Ridge Model:", ridge_cv_mae, "\n")
## Cross-Validation MAE for Ridge Model: 0.07678223
cat("Cross-Validation R2 for Ridge Model:", ridge_cv_rsqa, "\n")
## Cross-Validation R2 for Ridge Model: 0.689611

```

Nota. Resultados de la validación cruzada para el modelo ridge.

El RMSE promedio en la validación cruzada fue de 0.1023, mientras que el error estándar residual (RSE) del modelo completo sin validación cruzada fue de 0.0999. Esta diferencia no es muy significativa, lo cual sugiere que el modelo Ridge tiene un rendimiento consistente en los diferentes conjuntos de entrenamiento y prueba, lo cual es un buen indicio de que el modelo no presenta un sobreajuste grave.

Por otro lado, el valor de R^2 en el modelo original sin validación cruzada fue de 0.7331, mientras que el R^2 promedio durante la validación cruzada fue de 0.6896. Esta disminución en R^2 durante la validación cruzada indica que el modelo está explicando algo menos de la varianza de

los datos cuando se valida en distintos subconjuntos. Sin embargo, la disminución no es drástica, lo cual sugiere que el modelo generaliza adecuadamente y las variaciones entre los conjuntos de entrenamiento y prueba no son significativas.

Por lo tanto, se puede establecer que el modelo Ridge sigue siendo adecuado para describir la relación entre las variables predictoras y la degradación del bosque en el Parque Tinigua-Macarena.

Definición de un Indicador de Riesgo de Deforestación para el PNN Chibiriquete

El desarrollo del indicador de riesgo de deforestación (IRD) para el Parque Nacional Natural Chibiriquete se apoya en los resultados obtenidos del Modelo Aditivo Generalizado (GAM). Este modelo se ha utilizado debido a su capacidad para capturar relaciones no lineales entre las variables de riesgo y la degradación del bosque, proporcionando así una representación detallada del comportamiento de cada factor en diferentes contextos espaciales y temporales. A partir de los coeficientes estimados por el modelo GAM, se busca construir un indicador que refleje de manera precisa la contribución de cada factor al riesgo de deforestación. La propuesta descrita a continuación integra tanto la magnitud del efecto de cada variable como su significancia estadística, para asegurar una evaluación equilibrada y robusta de los factores que impulsan la degradación del bosque en el Parque Chibiriquete.

Para calcular el Indicador de Riesgo de Deforestación (IRD), se utiliza la siguiente fórmula:

- $$IRD = w1 \cdot \text{scaled_Cabezas_Ganado} + w2 \cdot \text{scaled_Cicatrices_Quema} + w3 \cdot \text{scaled_Puntos_Calor}$$

Donde cada peso ($w1$, $w2$, $w3$) representa la importancia relativa de cada variable, calculada según la metodología descrita a continuación.

Para ponderar los pesos del indicador de riesgo de deforestación (IRD) en el Parque Nacional Natural Chibiriquete, se propone un enfoque basado en la combinación del F-value y el p-value de cada variable involucrada. Esta metodología permite capturar tanto la magnitud del efecto (representada por el F-value) como la certeza de que dicho efecto es estadísticamente robusto (medido mediante el p-value). Con esta estrategia se busca balancear entre la cantidad de

varianza explicada por cada factor y la certeza de la influencia de dicho factor, para generar una representación justa y precisa de la importancia relativa de cada variable en la deforestación.

Propuesta de Ponderación de Pesos Usando F-value y p-value

Calcular un factor de importancia para cada variable: El objetivo es combinar el F-value (que mide la magnitud del efecto) y el p-value (que mide la certeza del efecto) para cada una de las variables independientes. A continuación, se describe cómo se puede llevar a cabo este proceso:

Normalizar los F-values: Para comenzar, se normalizan los F-values con el fin de garantizar que estén en un rango comparable. Para cada variable, la normalización se calcula de la siguiente manera:

$$F_{normalizado_i} = \frac{F_i}{\sum_1^n F}$$

Donde F_i es el F-value de cada variable y $\sum_1^n F$ es la suma de todos los F-values de las variables consideradas.

Asignar una penalización basada en el p-value: Para reflejar la certeza estadística de cada variable, se define una penalización para los pesos basada en el p-value. Esta penalización se puede representar mediante una función inversa del p-value, es decir:

$$P_{factor_i} = \frac{1}{P_i}$$

Donde P_i es el p-value de cada variable. Para evitar que los p-values muy pequeños generen pesos excesivamente grandes, se puede establecer un límite inferior razonable (ej. $P_i \geq 0.001$).

Calcular un peso combinado para cada variable: El peso combinado se obtiene multiplicando los F-values normalizados por los factores derivados de los p-values:

$$W_i = F_{normalizado_i} \times P_{factor_i}$$

Este enfoque asegura que tanto la magnitud del efecto (F-value) como la certeza de la significancia (p-value) se reflejen adecuadamente en el peso final de cada variable.

Normalización final de los pesos: Finalmente, los pesos combinados se normalizan para que sumen 1, lo cual facilita su interpretación y aplicación al cálculo del IRD:

$$W_{normalizado_i} = \frac{W_i}{\sum_1^n W}$$

De esta manera, se obtienen los pesos finales normalizados para cada una de las variables en el cálculo del IRD.

Determinación de la Fórmula para el IRD

Considerando el resumen del modelo GAM obtenido, se extraen los valores para los F-values y p-values para las variables de "scaled_Puntos_Calor", "scaled_Cicatrices_Quema" y "scaled_Cabezas_Ganado":

F-values:

- $F_{scaled_puntos_calor} = 1.371$
- $F_{scaled_cicatrices_quema} = 7.038$
- $F_{scaled_cabezas_ganado} = 3.447$

p-values:

- $p_{scaled_puntos_calor} = 0.24397$
- $p_{scaled_cicatrices_quema} = 0.00594$
- $p_{scaled_cabezas_ganado} = 0.00159$

Paso a Paso del Cálculo de Pesos

Normalización de los F-values.

- Suma de los F-values: $1.371 + 7.038 + 3.447 = 11.856$
- $F_{normalizado_scaled_puntos_calor} = \frac{1.371}{11.856} = 0.1156$
- $F_{normalizado_scaled_cicatrices_quemar} = \frac{7.038}{11.856} = 0.5934$
- $F_{normalizado_scaled_cabezas_ganado} = \frac{3.447}{11.856} = 0.2908$

Calcular el factor inverso del p-value:

- $P_{factor_scaled_puntos_calor} = \frac{1}{0.24397} = 4.10$
- $P_{factor_scaled_cicatrices_quema} = \frac{1}{0.00594} = 168.35$
- $P_{factor_scaled_cabezas_ganado} = \frac{1}{0.00159} = 628.93$

Calcular pesos combinados:

- $W_{scaled_puntos_calor} = 0.1156 \times 4.10 = 0.474$
- $W_{scaled_cicatrices_quema} = 0.5934 \times 168.35 = 99.93$
- $W_{scaled_puntos_calor} = 0.2908 \times 628.93 = 182.91$

Normalización de los pesos combinados:

- Suma de los pesos combinados: $0.474 + 99.93 + 182.91 = 283.314$
- $W_{final_scaled_puntos_calor} = \frac{0.474}{283.314} = 0.0017$
- $W_{final_scaled_cicatrices_quema} = \frac{99.93}{283.314} = 0.3525$
- $W_{final_scaled_cabezas_ganado} = \frac{182.91}{283.314} = 0.6458$

Interpretación de los Pesos:

Cabezas de ganado tiene el mayor peso (0.6458), lo cual refleja tanto su significancia estadística alta como su magnitud de efecto considerable sobre la degradación del bosque.

Cicatrices de quema también tiene un peso relevante (0.3525), debido a su efecto significativo.

Puntos de calor tiene un peso muy bajo (0.0017), reflejando que, aunque contribuye, su impacto es menor tanto en magnitud como en certeza.

Finalmente, se expresa la fórmula que sirve para el cálculo del indicador:

- $IRD = 0.6458 \cdot scaled_Cabezas_Ganado + 0.3525 \cdot$

$$scaled_Cicatrices_Quema + 0.0017 \cdot scaled_Puntos_Calor$$

Interpretación:

- Un IRD cercano a 1 indicaría un mayor riesgo de deforestación.
- Un IRD cercano a 0 indicaría un riesgo bajo de deforestación.

Ejemplos Aplicados

A continuación, se presentan dos ejemplos aplicados del cálculo del indicador de riesgo de deforestación (IRD) para el parque nacional natural serranía del Chibiriquete donde puede evidenciarse la influencia de los pesos para casos específicos de valores de las variables predictoras.

Tabla 2

Ejemplo Cálculo IRD parque Chibiriquete

	Ejemplo aplicado 1	Ejemplo aplicado 2
scaled_puntos_calor (s_pc)	0,8	0,2
scaled_cicatrices_quema (s_cq)	0,5	0,7
scaled_cabezas_ganado (s_cg)	0,2	0,9
IRD = (0.0017 * s_pc) + (0.3525 * s_cc) + (0.6458 * s_cg)	0,31	0,83

Interpretación	El IRD de 0.31 sugiere un riesgo bajo-moderado de deforestación. A pesar de la alta presencia de puntos de calor, la baja concentración de ganado y el valor medio de cicatrices de quema contribuyen a que el riesgo general no sea tan alto. Esto puede indicar que los puntos de calor están presentes por fenómenos como los incendios forestales o quemadas controladas, pero no necesariamente como acción para expandir la frontera agropecuaria.	El valor del IRD de 0.83 indica un riesgo alto de deforestación. En este escenario, la combinación de una alta concentración de ganado y una presencia significativa de cicatrices de quema son los principales impulsores del riesgo, incluso con una baja presencia de puntos de calor. Esto sugiere que la deforestación podría estar ocurriendo debido a la expansión de la frontera agrícola para la ganadería, y que la presencia de cicatrices de quema podría estar relacionada con la preparación del terreno para este fin.
----------------	--	---

Nota. Ejemplos aplicados del cálculo del indicador de riesgo de deforestación para el parque nacional natural serranía del chibiriquete.

Ventajas de Este Enfoque:

Balance entre magnitud y significancia: al combinar F-values y p-values, se logra un equilibrio entre la cantidad de varianza explicada y la confiabilidad del efecto.

Flexibilidad: Este método permite ajustar la importancia relativa de las variables dependiendo del contexto y del nivel de incertidumbre aceptable.

Esta combinación asegura que los pesos reflejen una representación justa de la importancia relativa de cada variable en el indicador, garantizando que tanto los efectos grandes como los confiables se tomen en cuenta en el cálculo del IRD.

Definición de un Indicador de Riesgo de Deforestación para el PNN Tinigua-Macarena

El desarrollo del indicador de riesgo de deforestación (IRD) para el Parque Nacional Natural Tinigua-Macarena se basa en el modelo de regresión Ridge. Este enfoque es particularmente adecuado debido a su capacidad para gestionar problemas de colinealidad entre las variables predictoras y generar coeficientes regularizados que evitan el sobreajuste del modelo. A partir de los coeficientes estimados con Ridge, se construye un indicador que pondera adecuadamente la contribución de cada factor al riesgo de deforestación.

Los coeficientes del modelo Ridge reflejan la relación entre cada variable (puntos de calor, cicatrices de quema y cabezas de ganado) y la degradación del bosque. Dado que estos coeficientes ya han sido regularizados para reducir la colinealidad y mejorar la estabilidad del modelo, se pueden usar como pesos en la fórmula del indicador de riesgo de deforestación. A continuación, se describe el proceso para asignar estos pesos y construir el IRD para el Parque Tinigua-Macarena.

Adaptación del Método de Pesos Propuesto

En el caso del Parque Chibiriquete, los pesos se derivaron mediante la normalización de los F-values y los p-values de cada variable. Sin embargo, en el contexto del modelo Ridge no se disponen de estos valores de significancia tradicionales, pero los coeficientes regularizados proporcionan una indicación clara de la importancia relativa de cada predictor.

Determinación de la Fórmula para el IRD

Para calcular el Indicador de Riesgo de Deforestación (IRD), se utiliza la siguiente fórmula:

- $$IRD = w1 \cdot \text{scaled_Cabezas_Ganado} + w2 \cdot \text{scaled_Cicatrices_Quema} + w3 \cdot \text{scaled_Puntos_Calor}$$

Donde cada peso (w_1 , w_2 , w_3) representa la importancia relativa de cada variable, calculada según la metodología descrita a continuación.

Asignar los Pesos a partir de los Coeficientes Ridge

Se utilizan los coeficientes regularizados del modelo Ridge como referencia para asignar pesos a cada variable. Si los coeficientes estimados para el modelo Ridge son:

- Coeficiente *scaled_Puntos_Calor*: 0.072
- Coeficiente *scaled_Cicatrices_Quema*: 0.003
- Coeficiente *scaled_Cabezas_Ganado*: 0.209

Estos coeficientes sirven como base para ponderar la contribución de cada variable en el indicador. En este caso, la variable "cabezas de ganado" tiene un mayor peso, seguida por "puntos de calor", mientras que "cicatrices de quema" tiene un peso mucho menor.

Normalización de los Coeficientes

Para obtener pesos proporcionales y que sumen 1, se normalizan los coeficientes dividiéndolos por la suma de todos los coeficientes:

- $Suma\ total = 0.072 + 0.003 + 0.209 = 0.284$
- $W1\ scaled_Cabezas_Ganado = 0.209 / 0.284 = 0.736$
- $W2\ scaled_Cicatrices_Quema = 0.003 / 0.284 = 0.011$
- $W3\ scaled_Puntos_Calor = 0.072 / 0.284 = 0.253$

Estos valores normalizados se utilizarán como los pesos en el cálculo del indicador. La fórmula para el indicador es, por lo tanto:

- $IRD = 0.253 \times scaled_Puntos_Calor + 0.011 \times scaled_Cicatrices_Quema + 0.736 \times scaled_Cabezas_Ganado$

El valor de IRD será una combinación lineal de las variables escaladas, ponderada según la importancia relativa de cada una según el modelo Ridge.

Interpretación:

- Un IRD cercano a 1 indicaría un mayor riesgo de deforestación.
- Un IRD cercano a 0 indicaría un riesgo bajo de deforestación.

Ejemplos Aplicados

Tabla 3

Ejemplo Cálculo IRD parque Tinigua-Macarena

	Ejemplo aplicado 1	Ejemplo aplicado 2
scaled_puntos_calor (s_pc)	0,9	0,1
scaled_cicatrices_quema (s_cq)	0,4	0,8
scaled_cabezas_ganado (s_cg)	0,3	0,7
IRD = (0.253 * s_pc) + (0.011 * s_cq) + (0.736 * s_cg)	0,45	0,55
Interpretación	El IRD de 0.45 sugiere un riesgo moderado de deforestación. La alta presencia de puntos de calor es el principal factor que contribuye al riesgo en este caso, aunque la concentración de ganado y la presencia de cicatrices de quema son relativamente bajas. Es crucial investigar las causas de los puntos de calor, ya que podrían indicar incendios forestales o quemas intencionales que requieren una intervención oportuna	El IRD de 0.550 indica un riesgo moderado de deforestación. En este escenario, la alta concentración de ganado y la presencia significativa de cicatrices de quema son los principales impulsores del riesgo, a pesar de la baja presencia de puntos de calor. La combinación de estos factores sugiere que la deforestación podría estar ocurriendo de manera más gradual, posiblemente relacionada con la expansión de la ganadería y la práctica de quemas controladas para la preparación de pastos.

Nota. Ejemplos aplicados del cálculo del indicador de riesgo de deforestación para el parque nacional natural Tinigua-Macarena.

Conclusiones

En el presente trabajo se ha examinado el complejo fenómeno de la deforestación en la Amazonía colombiana, centrado específicamente en los Parques Nacionales Naturales Chibiriquete, Tinigua, y Sierra de la Macarena. A través de un análisis espaciotemporal de factores de riesgo como puntos de calor, cicatrices de quema, y la explotación ganadera, se identificaron patrones críticos que contribuyen a la degradación forestal en estas áreas protegidas. A continuación, se presentan las principales conclusiones del estudio:

Variabilidad temporal y espacial de los factores de riesgo: Los análisis realizados evidencian una marcada variabilidad temporal y espacial en los factores de riesgo de deforestación. Los puntos de calor y las cicatrices de quema mostraron incrementos significativos en los años 2018, 2020, y 2022, lo cual subraya la relevancia de eventos específicos que ejercen una presión considerable sobre el ecosistema. Estos patrones sugieren que existen períodos críticos donde los esfuerzos de conservación deben intensificarse para prevenir la degradación irreversible del bosque.

Impacto de la ganadería extensiva: El crecimiento sostenido del número de cabezas de ganado en los municipios adyacentes a los parques emerge como uno de los principales factores de presión sobre los ecosistemas protegidos. Municipios como San Vicente del Caguán, Vistahermosa, y La Macarena mostraron un crecimiento sustancial en la actividad ganadera, directamente vinculada a la conversión de tierras forestales en pastizales. Este proceso de praderización representa una de las principales amenazas para la integridad de estos parques y subraya la necesidad de implementar prácticas de ganadería sostenible que mitiguen la deforestación.

Concentración de eventos extremos: Los boxplots de las distancias de puntos de calor y cicatrices de quema revelan la presencia de valores atípicos significativos en varios municipios. Estos valores sugieren que, aunque la mayoría de los eventos de quema y puntos de calor ocurren a distancias moderadas, existen episodios extremos que afectan áreas muy cercanas o muy alejadas de los parques, lo cual complica el manejo del riesgo de deforestación. Estos eventos extremos requieren estrategias de respuesta rápida y coordinada para mitigar su impacto.

Evaluación de resultados mediante validación cruzada: Los modelos GAM y Ridge se evaluaron utilizando técnicas de validación cruzada para explorar su capacidad de generalización y buscar reducir el sobreajuste. En el caso del modelo GAM aplicado al Parque Nacional Natural Chibiriquete, la validación cruzada reveló diferencias entre las métricas de desempeño en los conjuntos de entrenamiento y prueba, sugiriendo cierto nivel de sobreajuste. El ajuste del parámetro de suavidad (k) no logró traducirse en una mejora. Sin embargo, el conjunto de prueba aun así explica aproximadamente el 94% de la variabilidad de la degradación del bosque. Para el modelo Ridge del Parque Tinigua-Macarena, la validación cruzada indicó un nivel aceptable de generalización, con diferencias mínimas en las métricas de rendimiento entre la validación cruzada y el conjunto de prueba. Aunque se observó una ligera disminución en el R^2 , el modelo sigue siendo adecuado para su aplicación en nuevos datos, lo que sugiere una capacidad aceptable.

Modelado del indicador de riesgo de deforestación (IRD): El desarrollo del IRD ha permitido cuantificar y categorizar el riesgo de deforestación en los parques estudiados, considerando la influencia de los factores mencionados. Los modelos estadísticos empleados, como el GAM y Ridge, han demostrado ser herramientas efectivas para capturar la variabilidad temporal y espacial de los factores de riesgo. El IRD se perfila como una herramienta importante

para orientar la gestión de la conservación de manera más precisa, permitiendo la identificación de las áreas de mayor riesgo que requieren atención prioritaria.

Importancia de la vigilancia en tiempo casi real: El uso de plataformas como Google Earth Engine y Dynamic World ha sido crucial para acceder a datos en tiempo casi real sobre cambios en la cobertura de la tierra. La capacidad de detectar patrones de deforestación de manera oportuna es esencial para implementar estrategias de conservación efectivas. La actualización constante de los datos permite una mejor comprensión del fenómeno de la deforestación y posibilita una respuesta más eficiente por parte de las autoridades competentes.

Necesidad de políticas integradas de conservación: Los resultados obtenidos enfatizan la necesidad de formular políticas integradas de conservación que aborden tanto la prevención de incendios como la regulación de la ganadería y la gestión del uso del suelo. La colaboración entre diferentes actores, incluidas comunidades locales, ONGs, y autoridades gubernamentales, es esencial para garantizar la protección efectiva de estos valiosos ecosistemas. Además, la implementación de sistemas de alerta temprana y el fortalecimiento de la vigilancia en zonas de alto riesgo son acciones prioritarias para mitigar los impactos de la deforestación.

Recomendaciones

A partir de los resultados y conclusiones del presente trabajo de grado, se proponen una serie de recomendaciones destinadas a mejorar la comprensión y gestión del fenómeno de la deforestación en la Amazonía colombiana, así como para orientar futuras investigaciones en la región. Estas recomendaciones se fundamentan en la identificación de las limitaciones actuales y en la oportunidad de enriquecer el modelado de riesgo de deforestación a través de la inclusión de variables adicionales y del uso de nuevas tecnologías.

Ampliación del conjunto de variables a estudiar: Para una comprensión más completa del fenómeno de la deforestación, se recomienda ampliar el conjunto de variables incluidas en el análisis. Variables como la densidad de población, la distribución de cultivos lícitos e ilícitos, la ubicación de fincas ganaderas y la trazabilidad del ganado en la zona, las vías terciarias y trochas, la presencia de grupos armados ilegales, y la cobertura institucional del estado (incluyendo planes de desarrollo y proyectos de infraestructura) son todas factores que influyen significativamente en la deforestación. Aunque estos datos son fragmentarios o no están actualizados, su futura integración podría proporcionar una visión más holística y precisa del problema, permitiendo un mejor modelamiento y predicción del riesgo de deforestación.

Fortalecer el uso de tecnologías de monitoreo remoto: Se recomienda explorar la factibilidad de utilizar plataformas como Google Earth y Dynamic World para el estudio del crecimiento de vías terciarias y trochas en la región amazónica. Estas infraestructuras de transporte son un factor clave en la expansión de la frontera agrícola y, por ende, en la deforestación. El monitoreo continuo del crecimiento de estas vías a partir de imágenes satelitales podría permitir la detección temprana de nuevas rutas ilegales y la implementación de acciones de control antes de que se intensifique la degradación forestal.

Asimismo, se sugiere evaluar la posibilidad de incorporar Dynamic World como herramienta de clasificación de cobertura de tierra para el análisis de cambios en el uso del suelo. Dynamic World podría ser utilizado por organizaciones de protección ambiental y el gobierno para obtener información más actualizada sobre las coberturas de la tierra. Actualmente, la metodología Corine se utiliza para la determinación de coberturas y sus cambios, pero presenta un retraso considerable que dificulta el monitoreo oportuno del fenómeno. Dynamic World podría ayudar a superar esta limitación, proporcionando datos de alta resolución y en tiempo más cercano al real.

Promover la integración de datos fragmentarios: Los datos relacionados con la trazabilidad del ganado, el crecimiento de cultivos, y la existencia de trochas a menudo se encuentran en diferentes bases de datos con un grado de integración limitado. Es importante fortalecer la recolección de datos de campo y fomentar la colaboración entre instituciones para construir bases de datos integradas y actualizadas. Esta integración permitirá la generación de modelos más robustos que capten las relaciones entre diferentes factores de riesgo y la deforestación, mejorando la capacidad predictiva y la efectividad de las estrategias de conservación.

Evaluar el impacto del conflicto armado y la gobernanza local: La influencia del conflicto armado y la presencia del Estado, junto con sus planes de gobierno y gobernanza local, también representan variables de alto impacto en la deforestación. Es necesario estudiar con mayor detalle cómo estos factores afectan la gobernanza del territorio y la conservación de los recursos naturales. Se sugiere la realización de estudios interdisciplinarios que involucren ciencias sociales y ciencias ambientales, para comprender mejor cómo la dinámica de conflicto y la intervención del Estado moldean el uso del suelo y la deforestación en estas zonas protegidas.

Retroalimentación de expertos y stakeholders: Para asegurar la relevancia y aplicabilidad del IRD, se recomienda involucrar a expertos en conservación y a actores locales en un proceso de validación cualitativa. La retroalimentación de estos actores es esencial para ajustar el indicador a las realidades locales y garantizar que sea una herramienta útil para la gestión de la conservación.

Ajuste de los modelos y evaluación de validación cruzada: Con base en los resultados obtenidos de las validaciones cruzadas realizadas para los modelos GAM y Ridge, se recomienda considerar ajustes adicionales en los modelos, como el ajuste del parámetro de suavidad (k) en los modelos GAM, con el fin de optimizar el balance entre precisión y generalización. Las validaciones cruzadas mostraron diferencias entre los resultados en los conjuntos de prueba y los resultados en la validación cruzada, lo cual sugiere un nivel de sobreajuste. Aunque dichas diferencias no fueron significativas, es importante mantener un enfoque proactivo para refinar los modelos, lo cual podría incluir la eliminación de variables poco significativas o el incremento en la cantidad y calidad de los datos disponibles para el entrenamiento de los modelos. Estos esfuerzos ayudarán a reducir posibles sesgos y mejorar la capacidad del modelo para generalizar a nuevos datos.

Referencias Bibliográficas

- León-Parra, A., Cortés-Gómez, A., Gutiérrez-Chacón, C., Solano-Gutiérrez, C. L., Fagua-González, G., López-Franco, M., & Franco, O. (2023). *Parques Nacionales Cómo Vamos - PNCV: “Los Parques Nacionales Naturales colombianos: amenazas y oportunidades para su gestión. Informe 2023”*. <https://parquescomovamos.com>
- Grigorev Alexey. (2021). *Machine Learning Bookcamp*. Manning Publications Co.
- Indepaz. (2024, febrero 22). *Ganadería y Deforestación en la Amazonía*.
<https://indepaz.org.co/ganaderia-y-deforestacion-en-la-amazonia/>
- Inés González, G., & Gómez, L. A. (2020). Vía al bosque amazónico ¿para qué y cómo? Preguntas clave. *Revista Colombia Amazónica*, 12, 81-90.
- James, Gareth., Witten, Daniela., Hastie, Trevor., & Tibshirani, Robert. (2017). *An introduction to statistical learning : with applications in R*. Springer : Springer Science+Business Media.
- Marina, L., Cárdenas, M., Gonzalo Murcia García, U., & Arias, J. E. (2017). El corazón de la amazonía colombiana (Contexto del Parque Nacional Natural Serranía de Chiribiquete). *Revista Colombia Amazónica*, 2017, 235-252.
- Sistema de información territorial amazonía colombiana. (2022). *La Amazonia Región Amazónica colombiana*. SIAT-AC. <https://siatac.co/la-amazonia-colombiana/>